Dual Labour Markets.
A Survey

Rüdiger Wapler*

Abstract

Whereas the standard modern theories of unemployment were developed in the context of a single sector labour market, this paper presents a survey of how these theories can be integrated into a dual labour market setting. This approach dichotomises the labour market into two sectors, a primary with high wages and high job security, and a secondary sector with lower wages and higher labour turnover. The survey consists of three main parts. Part one applies standard labour market theories to dual labour markets. Part two extends this setup by introducing uncertain product demand. Finally, in part three, open economies are discussed. It is shown that dual labour market theory is an important contribution to understanding unemployment and therefore a useful extension of standard labour economics.

JEL classification: J21, J23, J24, J31, J41, J51, J62, J64

*University of Tübingen, Dept. of Economic Theory (Prof. Dr. Manfred Stadler), Mohlstr. 36, D - 72074 Tübingen, Germany; E-mail: ruediger.wapler@uni-tuebingen.de

Contents

List of Figures

List of Variables Used

a	Elasticity of the revenue function with respect to labour
d	Detection rate
e	Effort level
e_h	"High" effort level
e_l	"Low" effort level
$g(\theta)$	Inverse matching function
$h(e)$	Effort function
j	Index: $j =$primary (skilled) or secondary (unskilled)
k^1	Unit costs in the primary sector
k^2	Unit costs in the secondary sector
m	matching function
n	Probability that a firm is inspected if it is complying to the minimum wage laws
p	Output price
p_1	Price of the primary sector good
p_2	Price of the secondary sector good
p^H	Price of the output good when product demand is high
$\tilde{p}$	Macroeconomic price index
q	Separation rate
r	Interest rate
s	Shirking variable
t	Time index
u	Utility
w	Wage rate
w_m	Minimum wage
$\bar{w}$	Output produced by an agent in either sector of the economy

$\underline{w}$	Minimum wage at which secondary firms can still find workers
w_1	Wage in the primary sector
w_1^A	Wages set by Nash arbitration in the primary sector
w_1^U	Wages set by unions in the primary sector
w_2	Wage in the secondary sector
w_2^m	Secondary wage when there are binding minimum wages
x_1	Primary good
x_1^i	Output by a firm i in the primary sector of the economy
x_2	Secondary good
x_2^i	Output by a firm i in the secondary sector of the economy
z	Fraction of workers who are skilled
D	Firing costs
$F(\bullet)$	Function
J	Expected present discounted value to a firm of a job
K	Total number of jobs available in the economy
L	Labour input
$\hat{L}$	Effective total labour in the primary sector
L_1	Number of workers employed in the primary workforce
L_1^m	Primary labour demand when minimum wages are binding
$\tilde{L}_1$	Number of primary work contracts offered at the beginning of the period
L_1^i	Labour input of firm i in the primary sector
L_1^A	Number of primary workers with a Nash-arbitrator
L_1^U	Number of primary workers with union bargaining
L_2	Number of workers employed in the secondary workforce
L_2^i	Labour input by firm i in the secondary sector
L_2^m	Secondary labour demand when minimum wages are binding
N	Number of agents in the economy
N_1	Number of agents in the primary sector
N_1^A	Number of agents in the primary sector with Nash arbitration
N_1^U	Number of agents in the primary sector when there is union bargaining
N_2	Size of the secondary sector
NSC	No-shirking condition
P	Penalty due if firms are caught paying below the minimum wage level
PMC	Product market condition
VMP	Value of marginal product

R	Revenue of a firm
U	Unemployment
V	Expected present value of lifetime utility
V_1	Expected present value of lifetime utility for a worker in the primary sector
V_1^N	Expected present value of lifetime utility in the primary sector for a non-shirking worker
V_1^S	Expected present value of lifetime utility for a shirking worker in the primary sector
V_2	Expected present value of lifetime utility for a worker in the secondary sector
W	Nash-maximand
X_1	Total output in the primary sector
X_2	Total output in the secondary sector
α	Instantaneous utility from shirking
β	Relative productivity of secondary compared to primary workers
θ	Probability that a firm finds a suitable worker
λ	Probability of a primary worker being retained in the next period
μ	Total derivative of unemployment with respect to the probability of finding a suitable worker
ν	Vacancy rate
π	Profit
ϖ	Total derivative of equilibrium unemployment with respect to the probability of finding a suitable worker
ρ	Subjective discount rate
ϱ	Relative price level of primary to secondary goods
υ	Unemployment rate
υ^A	Unemployment rate with Nash-arbitrator
υ^U	Unemployment rate with union bargaining
ϕ	Probability that product demand is high
ψ	Probability of finding a job in the primary sector
ω	Ratio of primary to secondary wages
Θ	Indirect utility
Λ	Ratio of primary to secondary workers
Υ	Vacancies
Ψ	Value of a vacancy

Chapter 1

Introduction

Unemployment has been a growing problem for many years now, especially in Europe, with persistently high unemployment rates. As SOLOW stated in 1986, "If macroeconomics is good for anything, it ought to be able to understand and explain this fact" (SOLOW 1986, S23). However, until the 1980's neoclassical labour market economics assumed perfect labour markets and treated them in the same way as any other type of market. This meant that, at least in the long run, a perfectly flexible wage ensured that unemployment could not occur.

Needless to say, a vast number of more recent theories now exist, which all try to explain firstly, why wages deviate from their competitive level, causing unemployment and secondly, why real wages remain relatively constant over time while at the same time there are large variations in employment levels. Up to now, no single general theory of unemployment has emerged. Instead, many theories co-exist, each being able to partially explain some of the causes of unemployment. All the theories have in common that non-Walrasian features of the labour market prevent it from clearing, in other words, that non-price rationing occurs for at least some jobs.

Whereas the standard versions of these theories were developed for a single sector labour market[1], the aim of this paper is to present a survey of how these theories can be integrated into a segmented or dual labour market. This theory analyses

[1] See for example, ROMER 1996, for an overview of theories of unemployment in a single sector labour market.

reasons for the common finding that while job characteristics may differ considerably among firms, the variations tend to be systematically related to each other. For example, higher wages, higher productivity, higher capital intensity, higher value-added and lower turnover rates are all correlated with each other. Further, wage differentials between broad sectors of the economy are fairly stable over time. While it is probably more accurate to think of a continuous pattern of segmentation, the dual labour market approach dichotomises the job market into a "primary" labour market, which fulfills the above characteristics, and a "secondary" tier, where all of the opposite is true. Essentially, wages are assumed more to be attached to jobs rather than workers, with all "good" jobs in one sector. Thus, if one occupation in an industry is highly paid, then all other occupations in that industry will be paid above average as well. Of course it is possible to argue that those sectors paying above average wages are simply compensating for undesirable working conditions or are hiring higher quality workers. However, this would still not explain why wages are higher for all occupations within one sector, e.g. why should the working conditions for a secretary in the mining industry be harsher than the working conditions for a secretary in other industries. This leads to the conclusion, that the wage differences are more likely to stem from competitive failures in the labour market and that emphasis needs to be placed on the differences between sectors of the economy.

Dual labour market theory dates back to work by HARRIS, TODARO (1970), who originally described the labour market in Africa as being composed of a rural agricultural and an urban manufacturing orientated sector as well as work by DOERINGER, PIORE (1971), who analysed the effects of poverty and discrimination in a dual sector setting. Although the concept of dual labour markets gained popularity during the early seventies, it fell rapidly into disrepute following criticisms by WACHTER (1974) and CAIN (1976), who argued that dual labour market theory was not compatible with neoclassical economics. However, since then, due to work by for example MCDONALD, SOLOW (1985) and BULOW, SUMMERS (1986) who applied the theory to industrialised countries within a stringent neoclassical theoretical framework, the theory has again gained importance leading DICKENS, LANG (1988) to talk of the reemergence of dual labour market theory.

One of the main differences between standard neoclassical and dual labour market theory is the nature of the steady state equilibrium which is characterised by two wage levels in equilibrium. As jobs in the primary sector are associated with higher utility than those in the secondary sector, the dual labour market is characterised by an excess supply of workers seeking jobs in the primary sector, i.e. access to some sectors is subject to non-price rationing. Dual labour market theory must thus be able to explain how and why in equilibrium firms pay different wages and offer different kinds of jobs and how this excess supply of workers for primary jobs can persist even if the market for secondary labour is perfectly competitive as is the case in most of the models presented in this survey. In this sense, dual labour market theory is just as much a theory of wage determination and of the allocation of workers to jobs with different wages as it is a theory of unemployment. The aim is thus to simultaneously analyse the microeconomic aspects of unemployment, i.e. who is unemployed, as well as the macroeconomic aspect of the actual level of unemployment.

Many early writers on labour market segmentation identified limited mobility amongst sectors as an important aspect of the theory. There was a hierarchy of sectors with access to the highest paying being the most difficult. More recently, the emphasis has shifted from mobility to queues. Evidence of excess supply in high wage jobs, wage differences unrelated to ability or job quality (and perhaps related to other characteristics not suggested by human capital theory) and the fact that workers in low wage jobs would prefer high-paying jobs, all provide evidence of queues.

This survey is organised as follows. Chapter 2 introduces different types of labour market imperfections. Chapter 2.1 presents a representative efficiency wage model, whereby firms can increase their profits by paying above market-clearing wages. In the subsequent chapters 2.2 and 2.3, different institutional settings, namely trade unions and government minimum wage legislation, prevent the wage from falling to clear the market. Chapter 2.4 analyses how labour market frictions in a matching model affect unemployment.

The theoretical framework developed in the above chapters is then extended in Chapter 3 by introducing goods market imperfections when product demand is uncertain. Chapter 4 applies dual labour market theory to open economies and Chapter 5 concludes.

Chapter 2

Dual Labour Markets with Different Types of Wage Determination

As stated in the introduction, the theory of dual labour markets is by its very nature a theory of the wage determination process, with the two labour market sectors being characterised by two different wage levels. As will be shown below, the wage level in the primary sector not only determines the unemployment level in this sector, it also explains how the dual sectors can coexist in a long-run steady state equilibrium.

This chapter presents several different types of wage determination processes, starting with the efficiency wage concept. There are many variants within this literature, but only the most common, the "no-shirking" model, will be discussed as the underlying idea is the same for all the concepts. Efficiency wages in general explain why it is rational for the profit maximising firm to pay wages above the market clearing level. This is different if wages are a result of bilateral negotiations between employers and unions, in which case the firm may not be able to choose its desired employment level given by its labour demand function, or if a binding minimum wage exists, whereby the firm is prevented from reducing the wage to its optimal level by legislation. The chapter ends with a matching model of frictional unemployment. Here it is search costs which lead to wages not falling to market-clearing levels.

2.1 Dual Labour Markets and Efficiency Wages

In standard neoclassical economics, the existence of unemployment and the implied failure of the labour market to clear can only result from the failure of wages to adjust to the point at which supply and demand for labour are equated. Efficiency wage theory explains why there are economic reasons for firms to pay wages that are in excess of market clearing wages.[1] The concept was pioneered by SOLOW (1979) and SALOP (1979) who assumed that agents obtain utility from the wage they earn but providing effort can be interpreted as costs employees incur which lead to disutility. For this reason, workers would prefer not to exert any effort, i.e. produce very little or in the extreme no output. The decision whether to provide effort or not can depend on such factors as the probability of being dismissed if one is caught not providing effort, or whether workers view their wage as being "fair", i.e. if they feel they are being justly treated relative to what other agents receive. If a worker is caught not providing any effort, the most a firm can do is dismiss him. However, if all workers receive the market clearing wage, then there would be no unemployment and someone dismissed from one firm could immediately find a job elsewhere. In this case, the worker would not suffer any income loss. If a firm now pays more than the market clearing wage, a worker who is dismissed for not providing any effort will suffer a real income loss because he[2] will not immediately be able to find another job paying this high a wage. A worker thus needs to weigh the cost of providing effort against the benefit of receiving a higher wage. The decision to pay higher wages is also rational for firms as due to this higher wage, workers provide effort and are therefore more productive than beforehand, i.e. unit labour costs are actually lower even if the nominal wage is higher. This also explains why it is impossible for an unemployed person to undermine this wage, as he could not credibly signal the firm that he would still provide effort at this lower wage. There are also other reasons why higher wages lead to higher productivity amongst workers. A higher wage will induce workers with higher reservation wages, in other words, generally

[1] See for example KATZ (1986) for an excellent survey of the efficiency wage literature.

[2] Throughout it will be assumed that workers can be either male or female. It is purely for simplification purposes that only the male form is explicitly named even though reference is made to both genders.

higher qualified individuals who have invested time and money into education, to apply for a job. Further, higher wages can also raise the loyalty amongst the workforce, thereby reducing the number of workers who voluntarily quit the firm each period, leading to lower unit labour costs.

It is efficiency wages that brought about a breakthrough in the study of dualism by introducing equilibrium analysis into a neoclassical framework to formally explain wage differences between sectors. In the "no-shirking" efficiency wage setup developed by SHAPIRO, STIGLITZ (1984), workers can only be imperfectly monitored as to whether they are providing effort or not. A worker detected shirking[3] is immediately dismissed.

Although there are a vast number of dual labour market models which use the "no-shirking" efficiency wage concept, the focus here will be on a simplified version of the model by BULOW, SUMMERS (1986), which is one of the most general. Apart from the introduction of a secondary sector, there are only minor differences between this model and the one originally formulated by SHAPIRO, STIGLITZ (1984). BULOW, SUMMERS (1986) assume constant returns to labour in production and product demand is less than completely elastic. As is standard in dual labour market theory, the "good jobs" are found in the primary sector, whereas the secondary sector is associated with lower wages and only casual attachments between workers and firms, leading to higher quit rates.

There are N identical infinitely lived agents, each supplying one unit of labour and producing $\bar{w}$ units of output in either sector. All agents are risk neutral and have homothetic preferences. Consumers have homothetic preferences and maximise their lifetime utility V which is given by

$$V = \int_t^\infty u(x_1, x_2 + \alpha s) exp^{-\rho t} dt \tag{2.1}$$

where u denotes utility, x_1 is the number of primary and x_2 the number of secondary goods consumed in time t and ρ is the subjective discount rate. Thus, consuming extra units of x_2 or shirking are perfect substitutes, with the instantaneous utility from shirking denoted by α and s correspondingly is a variable that equals one if the worker shirks and zero if he provides effort.

[3] The word "shirk" is widely accepted in the economic literature even though it has negative connotations. It simply postulates that workers are over some range happier working less than more hard.

There are no monitoring problems in the secondary sector which is assumed to be perfectly competitive. Therefore, workers receive a wage w_2 equal to their marginal productivity. Detecting shirkers in the primary sector is difficult, leading to firms paying efficiency wages. To keep the analysis as simple as possible, a constant hazard rate of detection d is assumed.[4] Further, a standard assumption in models of this type is that structural shifts in demand mean that in any period, some firms are forced to exit the primary sector whilst others, producing new products, enter the sector. Thus, if firms are assumed to be identical, there is a probability q in each period that the firm will be forced out of the market and therefore, that the employment relationship comes to an end.

The price p_2 of the secondary sector output is normalised to unity. Together with the homotheticity of the utility function, this implies that the price p_1 of the primary good can be written as

$$p_1 = F\left(\frac{x_1}{x_2}\right) = F\left(\frac{L_1}{L_2}\right), \qquad F' < 0 \tag{2.2}$$

where L_1 and L_2 are the number of workers employed in the primary and secondary sector respectively.

In equilibrium, primary firms will set wages equal to the value of their workers' marginal product. As workers produce the same amount of output in either sector, it must hold that

$$w_1 = p_1 w_2 \tag{2.3}$$

which, using equation (2.2) can be written as

$$w_1 = w_2 F\left(\frac{L_1}{L_2}\right) \tag{2.4}$$

Equation (2.4) defines the "product market condition" (PMC). Since workers produce the same amount of output in either sector, it shows the primary wage

[4] See for example ORDOVER, SHAPIRO (1984) for a model in which the optimal choice of supervision technology is explicitly analysed. Here it is assumed that the firm in a first stage has determined its optimal monitoring technology which in a second stage leads to the hazard detection rate d. BULOW, SUMMERS (1986) assume that even workers who are not shirking may falsely be labelled as such and are dismissed.

w_1 as a decreasing function of the relative size of the primary sector. If the primary sector increases in size, then c.p. this will lead to a fall in p_1, and, correspondingly, to a lower value of marginal product.

With unemployment and only steady-states being analysed, a worker will remain in the primary sector unless there is an exogenous shock. The present-value of holding a job in this sector, V_1, can therefore be written as

$$\begin{aligned} V_1 &= (w_1 - w_2) \int_0^\infty exp^{-(\rho+q+d)t} dt \\ &= \frac{w_1 - w_2}{\rho + q + d} \end{aligned}$$

Rearranging yields the "no-shirking condition" (NSC)

$$w_1 = w_2 + \alpha(\rho + q + d) \tag{2.5}$$

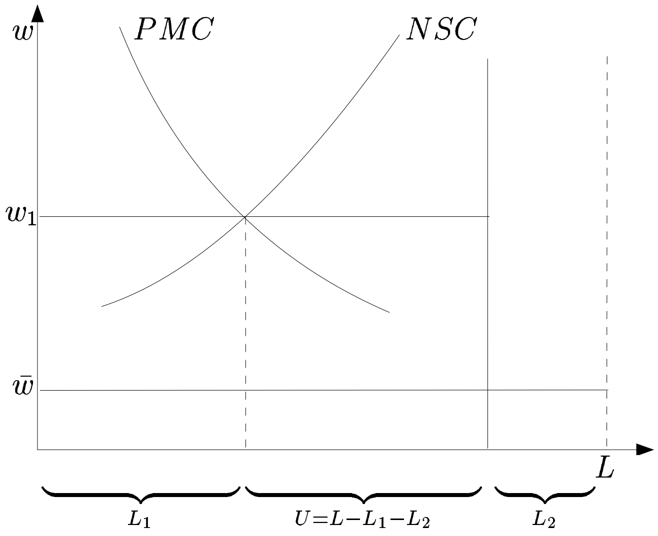
Using equation (2.5) it can be seen primary workers will only continue to provide effort if their wages increase as α, the utility from shirking increases. Similarly, if workers discount future income more, represented by an increase in ρ, then the utility loss and thus threat of being dismissed declines. Again, firms can compensate this effect by paying higher wages. Lastly, wages must also increase if the total turnover rate $q + d$ rises. This is because the value of maintaining a job is reduced if future turnover is more likely, as it will be easier to find a job if one is dismissed.

Before it is possible to analyse the steady state, particularly the sizes of the two sectors and thus the level of unemployment in an economy, it is necessary to make assumptions about how workers are divided amongst the sectors. In order to do this BULOW, SUMMERS (1986) integrate aspects of both voluntary and involuntary unemployment into their model by assuming that firms in the primary sector only hire individuals currently unemployed. Thus, there is "wait unemployment" as workers queue for primary jobs. This unemployment is voluntary in the sense that these workers could instantaneously obtain a job in the secondary sector, but it is involuntary in the sense that the unemployed would like to work at the wage paid in the primary sector.[5]

[5] Within the model it is not strictly possible to justify this assumption as it assumes that

The determination of the equilibrium of the "product market condition" (2.4) and the "no-shirking" condition (2.5) is shown in Figure 2.1. At the given level of un-

Figure 2.1: The Determination of Equilibrium Unemployment



employment, secondary sector workers must be indifferent to being unemployed.

Therefore, the utility gain from shirking, α, is equal to the present value of holding a job in the primary sector. This can be seen by inserting equation (2.3) into equation (2.5) and simplifying

$$\alpha = \frac{d(p_1 - 1)w_2}{\rho + q + d} \tag{2.6}$$

Finally, in a steady state, a worker must decide which sector to enter at the beginning of his career. Thus, the difference in instantaneous utility between holding a secondary sector job and remaining unemployed must equal the product

workers differ in their turnover probabilities or their tastes for primary sector work and these differences are unobservable to the firm. In that case, workers with low turnover probabilities or stronger preferences for primary work will opt for unemployment at the beginning of their careers, in the hope of later finding a job in the high-wage sector. Primary sector firms will then want to hire these individuals because the efficiency wages required to prevent workers from shirking are lower than they would be otherwise. Put differently, accepting a job in the secondary sector could be interpreted as a negative signal to primary sector employers, as it indicates a higher turnover probability. This theoretical reasoning is also supported empirically by CLARK, SUMMERS 1979, who find that secondary sector workers only rarely leave their jobs to become unemployed and seek primary work.

of the instantaneous probability of obtaining a primary sector job and the present value of such a job. Using equation (2.6) this can be written as

$$w_1 = \frac{(d+q)L_1}{N - L_1 - L_2} \cdot \frac{(p_1 - 1)w_2}{\rho + q + d} \tag{2.7}$$

From equations (2.4), (2.6) and (2.7) it is now possible to analyse how a positive productivity shock, i.e. an increase in worker output $\bar{w}$ affects unemployment. If consumers have very inelastic demands for the two goods, then an increase in the relative size of the primary sector and thus in the number of respective goods will only be consumed if the primary price level is greatly reduced. In this case, if consumers will always want to buy primary and secondary goods in the same relative amounts, then an increase in the number of primary goods must go along with a similar increase in the size of the secondary sector. Therefore, total unemployment will be reduced as both sectors grow proportionally.

If on the other hand relative prices do not change with changes in the sizes of the sectors of the economy, then the value of the marginal product increases and thus, from equation (2.5), primary wages will also increase. At the same time, this increase in productivity means that unit labour costs fall and therefore, that the primary sector will expand. However, with constant relative prices, the fraction of secondary goods consumed also remains constant. With increased labour productivity, it will now be possible to produce the same amount of secondary goods with fewer secondary workers so that unemployment will increase as a result. This effect is strengthened as the increase in the primary wage will lead some workers to exit the secondary and look for a job in the primary sector.

There are several policy implications implied by the model the most obvious of which is a high-wage sector subsidy which would lead to an increase in total output. However, this policy recommendation rests on the crucial assumption that monitoring is perfect in the secondary sector. Therefore, the outcome in this sector is always efficient and welfare improvements can only be achieved by increasing output in the primary sector. As Matusz (1994) points out, with unemployment in both sectors, wage or production subsidies will have two effects. On the one hand, the composition of the labour force will change as more workers are employed in the subsidised sector. On the other hand, there is also a level effect as total employment can either rise or decrease. In addition, even though primary firms cannot perfectly monitor their employees and there is only a positive probability of detecting shirkers, the costs of the monitoring technology are ignored. This means that the model does not explain the detection probability itself. Further, if there are no monitoring costs, then a profit maximising firm will fully monitor its workforce instead of paying higher wages.

Unemployment could also be decreased by reducing unemployment benefits to zero, as the only effect they have, is to reduce the value of the dismissal threat by firms, increasing the required primary wage as a result. Although such an extreme measure is unrealistic, in many industrialised countries there is an ongoing debate whether unemployment benefits should be lowered or not.

That workers are identical is critical in order to be able to calculate a steady state equilibrium. However, this assumption makes it impossible to endogenously determine how workers are distributed amongst the sectors. For this reason, it is especially critical to assume that primary firms only hire the currently unemployed. Although BULOW, SUMMERS (1986) justify this assumption using signalling considerations, these can only hold if workers have different rates of time preference or different job-turnover rates. This point is taken up in JONES (1987b) where workers are identical except for their turnover probabilities. He shows that with complete screening, only primary workers with lower turnover probabilities will choose to remain unemployed and therefore, that firms will only hire these workers, as the efficiency wage required to hire them is lower. However, if only incomplete screening is possible, then both secondary and primary workers may opt for voluntary unemployment waiting for a job in the primary sector. This shows that in a more fully developed model, it would be necessary to endogenise turnover rates, and by a similar reasoning, also to form a microeconomic foundation for the job separation rate.

Generally, efficiency wage models of the "no-shirking" type face the problem of explaining how they elicit effort from workers at the end of their careers. If a worker knows he is going to retire at the end of the period, then even if he shirks and is dismissed, he will not suffer any income (or utility) loss. A similar situation arises when product demand is uncertain, in which case, if demand is very low, it will be necessary for firms to fire workers, the implications of which will be discussed in more detail in Chapter 3. Also, the threat that a firm will immediately dismiss a worker detected shirking is only believable if workers do not have any firm-specific human capital. Added to this fact is that in general it will not be possible to immediately fire workers due to legal reasons. The longer it takes before a firm can dismiss a worker, the lower his income will be.

As stated at the outset, there are as many different efficiency wage concepts as there are economic reasons for firms to pay above market-clearing wages. Regardless of which efficiency wage theory is chosen, the theory suffers the disadvantage that the wage is solely governed by the effort function and is independent of factors such as capital intensity or profitability. Other theories of wage determination such as trade union bargaining theories or matching theories are able to overcome these difficulties. It is for this reason that in the remainder of this chapter the focus will be on these theories and on the role they play in determining unemployment in a dual labour market setting.

2.2 Dual Labour Markets and Trade Unions

One of the differences between the European – perhaps with the exception of the UK – and the US labour market which are often named is the influence of trade unions. In general, the primary aim of unions is to increase the wages of their members. Whereas in the simple Walrasian all workers were simply price-takers on the labour market, this is different if unions with a degree of bargaining strength are taken into account. Models on trade-union wage negotiations can generally be classed into two main categories. Firstly, there is the so-called "right-to-manage" approach. Here unions and employers only negotiate over the wage level and employers subsequently unilaterally determine the employment level given the bargained wage level. Secondly, in "efficient bargaining" models, unions and employers simultaneously determine the wage and employment level.

In his seminal work, Calvo (1978), discussed here in a simplified version, analyses two different setups. First, wage and employment levels are assumed under the premise that the unions have monopoly power in determining the wage rate. Secondly, a Nash-bargaining setup is chosen, in which unions and employers have equal influence on the wage level.

Important is that unions are only present in the primary sector. The secondary sector is again assumed to be perfectly competitive. For this reason, there is full employment in this sector and the wage is determined by the marginal productivity of labour which can be treated as being exogenously given by

$$\begin{aligned} X_2 &= F(L_2) \\ w_2 &= F'(L_2) \end{aligned} \tag{2.8}$$

with X_2 denoting total output, L_2 employment and w_2 the wage in the secondary sector. The price level in this sector is normalised to one.

Production in the primary sector is given by the concave function

$$X_1 = F(L_1) = L_1^a \qquad 0 < a < 1 \tag{2.9}$$

where X_1 are output and L_1 the (employed) labour force size in the primary sector and a is the elasticity of output with respect to labour. The workforce in the economy N is assumed to be homogeneous, so that it holds that

$$\begin{aligned} &N = N_1 + N_2 \\ &N_1 \geq L_1, \qquad N_2 = L_2 \end{aligned}$$

where N_1 is the size of the primary and N_2 that of the secondary workforce. The probability of finding a job in the primary sector ψ is given by

$$\psi = \frac{L_1}{N_1} \tag{2.10}$$

Finally, in equilibrium expected wages in both sectors must be equal

$$w_2 = w_1 \frac{L_1}{N_1} \tag{2.11}$$

The wage rate in the primary sector is determined by the interaction of unions and employers. Unions are assumed to derive utility u from higher wages for their members. Specifically, the unions aim is demand higher wages than those paid in the secondary sector so that their objective function is

$$u = L_1(w_1 - w_2) \tag{2.12}$$

From (2.12) it can be seen, that the lowest primary wage level possible is w_2 which is intuitively plausible, as any worker can always costlessly transfer to the secondary sector.[6]

The primary wage that employers can pay depends on the marginal productivity of primary labour

$$p_1 F'(L_1) = w_1 \tag{2.13}$$

with p_1 denoting the price level in this sector. Using the production function (2.9) it is possible to determine the wage elasticity of labour demand as

$$-\frac{\partial L_1}{\partial w_1}\frac{w_1}{L_1} = \frac{1}{1-a} \tag{2.14}$$

Unions on the other hand will set wages so as to maximise their utility. Denoting this wage by w_1^U means that the optimum wage is given by

$$\frac{\partial u}{\partial w_1^U} = \frac{\partial u}{\partial w_1^U} + \frac{\partial u}{\partial L_1^U}\frac{\partial L_1^U}{\partial w_1^U} = 0 \tag{2.15}$$

There are two opposing effects in (2.15). On the one hand, increases in the wage level also lead to an increase in utility for the unions as their members receive higher wages. On the other hand, the number of members will decline with higher wages, thereby reducing union utility. In the optimum these two effects need to be equalised. It therefore follows that

$$L_1^U = (w_1^U - w_2)\frac{\partial L_1^U}{\partial w_1^U}$$

which after rearranging and inserting (2.14) becomes

$$\begin{aligned} 1 &= \frac{\partial L_1^U}{\partial w_1^U}\frac{w_1^U}{L_1^U}\left(\frac{w_1^U - w_2}{w_1^U}\right) \\ w_1^U &= \frac{w_2}{a} \end{aligned} \tag{2.16}$$

Thus, it can be seen that the primary wage is determined only by the secondary wage and the technology used by primary firms and further, that it is impossible for the two wage levels to be equal. As a consequence of this higher primary wage and the fact that expected wages have to be identical in equilibrium, there

[6] A worker in the primary sector does not take into account that the secondary wage will fall if he switches sectors.

must be unemployment in the primary sector as some workers prefer to wait for a job to become free rather than working in the secondary sector at a lower wage. The exact number of agents in the primary sector when there is union wage bargaining, N_1^U, can be obtained by combining equations (2.9), (2.13) and (2.16) which yields

$$N_1^U = \left(\frac{a^{1+a}p_1}{w_2}\right)^{\frac{1}{1-a}} \tag{2.17}$$

As can be seen, improvements in technology or higher prices, both of which raise expected wages, make the primary sector more attractive thereby increasing the number of workers seeking jobs in this sector. As is intuitively to be expected, a higher secondary wage will lead some individuals looking for a primary job to stop seeking in this sector and take up a job in the secondary sector.

The actual unemployment rate associated with union bargaining, v^U, is simply derived by inserting the wage equation (2.16) into (2.11)

$$v^U = 1 - a \tag{2.18}$$

A higher elasticity of output with respect to labour, a, will lead more workers into the primary sector as firms expand their output. At the same time this reduces the size of the workforce in the secondary sector, which leads to an increase in the secondary wage diminishing the wage differential. Both of these effects reduce unemployment.

These results now need to be compared to those obtained if a Nash-arbitrator determines wages and employment in the primary sector. In this case, profits π are given by

$$\pi = F(L_1) - w_1 L_1 \tag{2.19}$$

Assuming that bargaining power is symmetrically divided between unions and employers, the Nash-maximand W is

$$W = (F(L_1) - w_1 L_1)^{\frac{1}{2}}(L_1(w_1 - w_2))^{\frac{1}{2}}$$

$$\frac{\partial W}{\partial L_1} = \frac{1}{2}(F(L_1) - w_1 L_1)^{-\frac{1}{2}}(F'(L_1) - w_1)(L_1(w_1 - w_2))^{\frac{1}{2}}$$
$$+ \frac{1}{2}(F(L_1) - w_1 L_1)^{\frac{1}{2}}(L_1(w_1 - w_2))^{-\frac{1}{2}}(w_1 - w_2) \stackrel{!}{=} 0$$

$$\frac{\partial W}{\partial w_1} = -\frac{1}{2}(F(L_1) - w_1 L_1)^{-\frac{1}{2}} L_1 (L_1(w_1 - w_2))^{\frac{1}{2}}$$
$$+ \frac{1}{2}(F(L_1) - w_1 L_1)^{\frac{1}{2}} (L_1(w_1 - w_2))^{-\frac{1}{2}} L_1 \overset{!}{=} 0$$

Combining these equations leads to

$$F'(L_1) = w_2 \tag{2.20}$$

$$u = \pi \tag{2.21}$$

Inserting (2.21) into (2.19) and denoting the labour force size and wage rate by the superscript A gives

$$\begin{aligned} 2u &= \left(L_1^A\right)^a - w_2 L_1 \\ 2L_1^A(w_1^A - w_2) &= \left(L_1^A\right)^a - w_2 L_1 \\ 2w_1^A &= \frac{w_2}{a} + w_2 \\ w_1^A &= w_2 \frac{1+a}{2a} \end{aligned} \tag{2.22}$$

As above, the size of the primary labour force N_1^A can now be found by combining equations (2.13) and (2.22) and inserting these into condition (2.11) to give

$$N_1^N = \frac{1}{2}\left(\frac{a^{1+a}p}{(1+a)^a w_2}\right)^{\frac{1}{1-a}} \tag{2.23}$$

and the corresponding unemplyoment rate

$$v^A = 1 - \frac{2a}{1+a} \tag{2.24}$$

Comparing the primary wage as determined by Nash-bargaining with that of trade unions, it can be seen that the former wage is higher. The reason for this is that with a Nash arbitrator, unions have lower bargaining strength which means that the disutility from higher wages but lower employment is weighted less. Firms do not suffer any loss in profits from higher primary wages, as, by (2.20), their optimal labour force size is only dependent on the secondary wage. Therefore, the higher wage is concomitant with higher unemployment and thus also a smaller primary labour force size as compared to the case of a monopoly union wage bargaining.

As shown above and irrespective of whether there is a trade union or a Nash arbitrator, by (2.16) and (2.22) wages in the two sectors can never be equal and are solely influenced by technological considerations. Since in a steady state expected wages need to be equal, this difference in wage rates automatically implies that there is unemployment in the high-wage sector. Therefore, all government measures must aim at equating expected wages in order to reduce unemployment. One such possible instrument would be to tax workers who switch from the secondary to the primary sector. That way, even though wages are higher in the primary sector, after tax wages are not, so that there is no incentive to search for a job and no unemployment in this sector. Similarly, a tax on primary output, lowering the price level in that sector, would of course also lower expected wages in the primary sector. A tax on primary wages (or conversely a wage subsidy on secondary wages) would only be effective if the union's objective function (2.12) were to only depend on pre-tax and not take-home wages, a fact which seems unrealistic. All of these policies have in common that a tax increase would lead to higher employment. However, it seems doubtful whether such policy measures are politically feasible.[7]

The setup presented here is extended by Burda (1992) to include unemployment benefits and labour force growth. As in the efficiency wage models discussed in Chapter 2.1, higher unemployment benefits make waiting for a job in the primary sector more attractive, leading to higher unemployment rates. A similar effect is derived for an increase in labour force growth. This increase in the labour supply reduces secondary wages, which means that expected wages are now higher in the primary sector, again leading to higher wait unemployment.

A dual labour market within the efficient bargaining literature is developed by McDonald, Solow (1985), who analyse the size of the wage differentials during a business cycle. As above, unions are only concerned with the utility of their employed members. This leads to a stable wage rate largely independent of overall economic conditions, but a widely fluctuating employment rate in the

[7] Although such arguments are similar to those often made when environmental taxes are discussed, the main difference is that the tax revenue from these taxes is to be used to either lower labour costs or the costs of preventing environmental pollution. In the model here, the mere fact that a tax is imposed no matter how the revenue is spent, can reduce unemployment to zero.

primary unionised sector. The wage rate in the secondary market is determined by market clearing conditions. Thus in this sector, shocks are absorbed more by wages than employment levels. For this reason, in an economic boom period, firms in the primary sector are willing to hire more workers (with no resulting change in the primary wage), leading many workers to leave the secondary sector hoping for a job in the primary sector.[8] This means that labour supply in the secondary sector is reduced driving up wages in this sector. Thus, in times of economic prosperity, wage differentials between the two sectors will decline and, by analogous reasoning, will increase during a recession.

All of the above models can be criticised in that they unrealistically assume that all employees need to be union members. Thus, it is impossible for any other worker to underbid the wage. In reality, such "closed shops" are forbidden in many countries. Further, even if there is wage bargaining, it is still possible that employers are also concerned about the efficiency of their employees. In this case, the production function needs to be rewritten as a function of effective labour as opposed to simply the quantity of labour employed.

The formulation of the union objective function is also extremely important as to the outcome of the model. Possible modifications would be that unions also take the national employment levels into account, with higher unemployment reducing union bargaining power. Similarly, it is of importance whether former employees, now unemployed but still union members, also determine union utility or not. Both of these adjustments are likely to lead to higher employment levels.

In general, higher welfare is reached in the efficient bargaining models, as in these in the resulting wage-employment equilibrium, it is impossible to make one party better off without harming the other. Thus, even though the government cannot directly intervene into the negotiations, one possible recommendation from the above would be to advise unions and employers to negotiate over both the wage and the employment level. In fact, this was done in the eighties in the Netherlands where employers, unions and the state all agreed on measures designed to increase

[8] In the original McDonald, Solow (1985) model some workers directly switch from the secondary to the primary sector, whilst others first switch to a transitional sector as the probability of finding a primary job is higher. This does not, however, alter the qualitative result that total secondary labour supply declines in a boom.

employment levels without having to drastically reduce wages. Such a measure is at the moment being tried in Germany as well, although it is far too early too tell how successful this will be. The range of government measures is of course completely different, if there is a national minimum wage which is binding. How this affects unemployment is the subject of the next section.

2.3 Dual Labour Markets and Minimum Wages

There is minimum wage legislation in many countries. In standard single sector models, minimum wages normally lead to higher unemployment rates amongst the low-skilled, and if one realistically assumes that coverage is not perfect, also lead to lower wages in the uncovered sector. These effects which decrease social welfare are of course partly compensated by the increase in welfare enjoyed by workers who remain employed and receive higher minimum wages, but it is unlikely that these positive effects outweigh the negative welfare effects. This chapter, using a model developed by Jones (1987a), analyses whether these results still hold in a dual labour market economy.

Efficiency wages are paid in both sectors of the economy. Due to differences in the nature of the jobs in the two sectors, monitoring is perfect in the secondary sector, whereas in the primary sector there is only the probability d that shirking is detected. For this reason, primary workers receive a wage premium in order to induce an efficient effort supply in equilibrium. The individual effort decision in the primary sector is determined by the wage rate and outside work opportunities. Thus, as opposed to the original Shapiro, Stiglitz (1984), model, it is not unemployment but wage differentials which serve as a worker discipline device.

Workers are homogeneous, live infinitely and are risk neutral so that instantaneous utility can be written as

$$u(w_j, e) = w_j - e \qquad i = 1, 2 \tag{2.25}$$

The expected discounted lifetime utility is

$$V = E\left[\int_0^\infty (w_j - e_t) exp^{-\rho t} dt\right] \tag{2.26}$$

where w_j is the wage in either the primary or secondary sector and e effort. For simplicity, it will be assumed that e either takes the value zero or some positive constant and that only workers providing effort are productive.

Firms in either sector are also risk neutral and face the same technology. There is perfect competition in the secondary sector, and because all workers must provide positive effort at all times, total output in this sector can be written as

$$X_2 = \max_{L_2^i} \sum_i F(L_2^i) \text{ s.t. } \sum_i L_2^i = L_2 = F(L_2) \tag{2.27}$$

where L_2^i are the number of secondary workers employed by firm i. Ignoring possible minimum wage laws momentarily, workers in this sector are paid a wage equal to their marginal product

$$w_2 = F'(L_2) \tag{2.28}$$

This means that the primary sector production function can be written as

$$x_1^i = F(L_1^i) \qquad F' > 0, F'' < 0 \tag{2.29}$$

The supervision technology is such that workers in this sector face the probability d of being detected shirking and fired as a consequence. Further, as above, each individual regardless of whether he shirks or not, faces the probability q of turnover out of this sector, whereas the probability of finding a job in this sector is given by ψ. Given these exogenous probabilities, in a steady state it holds that

$$qL_1 = \psi(N - L_1) \tag{2.30}$$

The probability of gaining access to the primary sector is equally likely no matter whether the agent is formerly in the secondary sector or unemployed.

Expected lifetime utility of a non-shirker employed in the primary sector V_1^N is

$$V_1^N = w_1 t - et + exp^{-\rho t}[qtV_2 + (1 - qt)V_1^N]$$

which can be rearranged to

$$\rho V_1^N = w_1 - e + q(V_2 - V_1^N) \tag{2.31}$$

Similar calculations for expected lifetime utility of a shirking worker and a worker currently in the secondary sector lead to

$$\rho V_1^S = w_1 - e + (d+q)(V_2 - V_1^S) \tag{2.32}$$

$$\rho V_2 = \frac{L_2}{N - L_1}(w_2 - e) + \psi(V_1 - V_2) \tag{2.33}$$

Equations (2.31) and (2.32) amount to a constant dividend plus an expected capital gain term due to a change in employment status. For a primary worker to choose $e > 0$ requires that $V_1^N \geq V_1^S$ which can be written as

$$w_1 \geq \rho V_2 + \frac{e(\rho + q + d)}{d} \tag{2.34}$$

Equation (2.34) represents the no-shirking condition (NSC) which requires that the expected loss from a transition out of the primary sector at least balances the instantaneous utility gain e that accrues to a shirker.

In equation (2.33) it is assumed that any worker leaving the primary sector has probability $L_2/(N - L_1)$ of secondary employment whereby he receives a net utility of $(w_2 - e)$ and V_1 corresponds to expected lifetime utility if employed in the primary sector given that the NSC defined in equation (2.34) holds. Using equations (2.30) – (2.33) it is possible to derive the steady state NSC as

$$w_1 \geq \frac{L_2}{N - L_1}(w_2 - e) + e + \frac{e}{d}\left(\frac{qN}{N - L_1} + \rho\right) \tag{2.35}$$

There are two possible types of equilibrium if minimum wage laws are neglected for the moment. If there is full equilibrium then it holds that

$$N = L_1 + L_2 \tag{2.36}$$

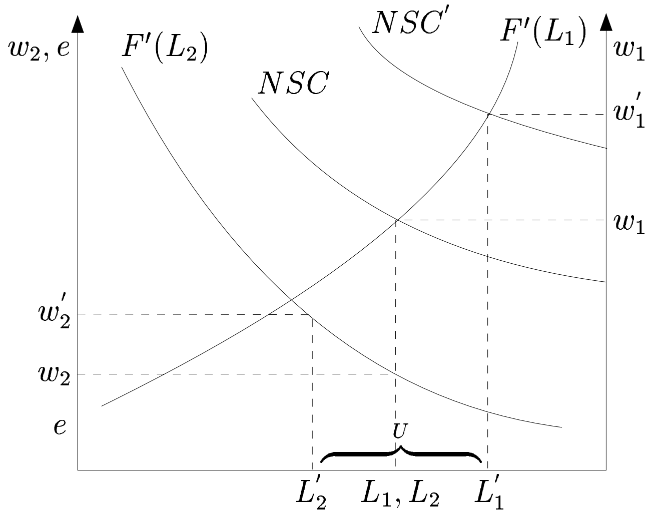
from which follows

$$F'(L_1) = w_1 = w_2 + \frac{e}{q}\left(\frac{qN}{N - L_1} + \rho\right) \tag{2.37}$$

$$F'(L_2) = w_2 \tag{2.38}$$

This case is shown in Figure 2.2. For this case to hold, the wage level in the secondary sector must be above or at least as high as the required effort level;

Figure 2.2: Dual Labour Market Equilibrium with Minimum Wages



if this is not the case, workers would prefer to remain unemployed. If the wage does fall below e, then equation (2.36) needs to be modified to

$$N > L_1 + L_2 \tag{2.39}$$

In this case, firms in the secondary sector would reduce their employment levels until the marginal productivity of labour is just equal to the required effort or productivity level e. This increase in the secondary wage also shifts the NSC upwards as it is dependent upon the opportunities outside of the primary sector. This is shown in Figure 2.2 by the secondary wage $w_2^{'}$, which leads to the new no-shirking condition $NSC^{'}$. From a welfare point of view, full employment could be reached with one wage rate applying in both sectors. However, at such a wage, no worker in the primary sector could credibly signal that he would not shirk.

Now minimum wage legislation is introduced. To keep the analysis simple, it is assumed that all firms comply to this law. The main case of interest and also the more realistic setting within a dual labour market setting is when the minimum wage laws are only binding in the secondary sector. If the minimum wage is denoted by w_m, then this condition can be expressed as $w_m > w_2$ where w_2 corresponds to the solution given by combining equations (2.36) and (2.37) when there is full employment, or by $w_2 = e$ when there is voluntary unemployment in the initial equilibrium. The condition that the minimum wage does not directly affect the primary sector can be determined by evaluating the no-shirking condition (2.35) at the wage rate w_m which yields

$$w_m \leq e + \frac{e}{d}\left(\frac{N - L_1}{N - L_1 - L_2}\right)\left(\frac{qN}{N - L_1} + \rho\right) \tag{2.40}$$

If condition (2.40) is violated, then there is no longer a dual structure in the economy. There would be no wage differential and the resulting unemployment acts as a worker discipline device.

In the range at which the minimum wage binds as described above, the dual labour market is given by

$$F'(L_2) = w_m \tag{2.41}$$

$$F'(L_1) = \frac{L_2}{N - L_1}(w_m - e) + e + \frac{e}{d}\left(\frac{qN}{N - L_1} + \rho\right) \tag{2.42}$$

In such a setup, the position of the NSC not only depends on the minimum wage w_m, but also on how much unemployment is associated with the minimum wage level, as both effects together determine the outside opportunities. The effect of changes in the minimum wage on the secondary sector are

$$\begin{aligned} \frac{\partial L_2}{\partial w_m} &= \left(\frac{\partial F/\partial L_2}{\partial L_2}\right)^{-1} \frac{\partial F/\partial L_2}{\partial w_m} \\ &= \frac{1}{F''(L_1)} < 0 \end{aligned} \tag{2.43}$$

and on the primary sector

$$\begin{aligned} \frac{\partial L_1}{\partial w_m} &= \left(\frac{\partial F/\partial L_1}{\partial L_1}\right)^{-1} \left(\frac{\partial F/\partial L_1}{\partial w_m} + \frac{\partial F/\partial L_1}{\partial L_2}\frac{\partial F/\partial L_2}{\partial w_m}\right) \\ &= \frac{\left(\frac{L_2}{N-L_1} + \left(\frac{w_m - e}{N-L_1}\right)\frac{\partial L_2}{\partial w_m}\right)}{F''(L_1) - \left(\frac{1}{N-L_1}\right)^2 \left(L_2(w_m - e) + \frac{eqN}{d}\right)} \end{aligned} \tag{2.44}$$

The denominator of (2.44) is always negative, however the sign on the numerator is ambiguous

$$\frac{\partial L_1}{\partial w_m} \gtrless 0 \text{ iff } \frac{L_2}{N - L_1} + \left(\frac{w_m - e}{N - L_1}\right)\frac{\partial L_2}{\partial w_m} \lessgtr 0$$

Using equations (2.41) and (2.43) this can be simplified to

$$\frac{\partial L_1}{\partial w_m} \gtrless 0 \text{ iff } L_2 < -\frac{F'(L_2) - e}{F''(L_2)} \lesseqgtr 0$$

which because $F''(L_2) < 0$ can be written as

$$\begin{aligned} \frac{\partial L_1}{\partial w_m} &\gtrless 0 \text{ iff } L_2 F''(L_2) > -F'(L_2) + e \\ &\gtrless 0 \text{ iff } L_2 F''(L_2) + F'(L_2) > e \end{aligned} \tag{2.45}$$

This ambiguity in employment effects in the primary sector arises because there are two opposing forces at work. On the one hand, a rise in the minimum wage will improve outside work opportunities and increase the required no-shirking primary wage rate. However, the higher minimum wage also means that there is now more unemployment in the secondary sector, which will reduce the required wage in the primary sector.

So far we have shown the effect on the size of the two respective labour forces. The total effect on unemployment can be determined using equations (2.43) and (2.44) which yields

$$\begin{aligned} \frac{\partial U}{\partial w_m} = \frac{-1}{F''(L_2)} &\left(1 + \frac{\frac{w_m - e}{N - L_1}}{F''(L_1) - \left(\frac{1}{N-L_1}\right)^2 \left(L_2(w_m - e) + \frac{eqN}{d}\right)} \right) \\ &- \frac{\frac{L_2}{N-L_1}}{F''(L_1) - \left(\frac{1}{N-L_1}\right)^2 \left(L_2(w_m - e) + \frac{eqN}{d}\right)} \end{aligned} \tag{2.46}$$

If (2.46) is evaluated at a voluntary unemployment equilibrium with $w_m = e$ then (2.46) simplifies to

$$\left.\frac{\partial U}{\partial w_m}\right|_{w_m = e} > 0$$

With initial voluntary unemployment, the marginal productivity of labour in the secondary sector given by equation (2.38) will be $F'(L_2) = e$. This means that equation (2.45) no longer holds, so that the rise in total unemployment is due to a drop in employment levels in both sectors simultaneously. Alternatively, if full employment with $N = L_1 + L_2$ is assumed, then rearranging and simplifying equation (2.46) shows that

$$\left.\frac{\partial U}{\partial w_m}\right|_{w_m = w_2 > e} > 0$$

In this case, the rise in total unemployment can be compromised of a fall in both sectoral employment levels as above, or alternatively, the rise in minimum wages is more than offset by the increase in total unemployment making it possible to decrease the primary wage rate w_1, which leads to higher employment levels in that sector. In the latter case, the minimum wage policy will be anti-egalitarian. Although the wage differential between the two sectors decreases, it is associated with an increase in unemployment levels in the secondary sector, i.e. it is possible that the low-skilled unemployment rate rises as a result.

The case when all firms comply to the minimum wage laws can be regarded as an extreme assumption. In reality empirical studies have shown at least for the United States, that non-compliance is substantial (see for example SELLEKAERTS, WELCH 1984). Here it is assumed that firms are inspected with probability n. The penalty P if they are found to be non-complying is that they must pay back the difference between the wages they paid and the minimum wages.

Within this setup, expected profits of a representative secondary sector firm i are

$$E[\pi_2^i] = F(L_2^i) - w_2^i - nP \tag{2.47}$$

where the penalty P is given by

$$\begin{array}{ll} 0 & \text{if } w_2^i \geq w_m \\ \frac{(w_m - w_2^i)L_2^i}{\psi} & \text{if } w_2^i < w_m \end{array} \tag{2.48}$$

where ψ was the job finding rate, so that its inverse $1/\psi$ denotes the average amount of time a worker stays in this sector before finding a job in the primary sector. The back wages in respect of the average as given by (2.48) are paid to all current employees. Thus, the effect that the labour force might have changed in size or composition during that time is ignored here.

The representative firm will then choose L_2^i and w_2^i to maximise its expected profits as given by equations (2.47) and (2.48) provided that it finds enough workers at this wage. Further, due to competition amongst secondary firms and labour supply conditions, it is assumed that firms cannot hire labour for less than $\underline{w}$. This means that the firm faces a constrained maximisation problem so that the Kuhn-Tucker conditions need to be applied

$$\max_{L_2^i, w_2^i} \pi_2^i = F(L_2^i) - w_2^i L_2^i - n\frac{(w_m - w_2^i)L_2^i}{\psi}$$

$$\text{s.t.} \quad w_2^i - \underline{w} \geq 0$$

$$w_2^i - w_m < 0$$

The typical Lagrangian function can be written as

$$\mathscr{L} = F(L_2^i) - w_2^i L_2^i - n\frac{(w_m - w_2^i)L_2^i}{\psi} + \lambda_1(w_2^i - \underline{w}) + \lambda_2(w_m - w_2^i)$$

$$\frac{\partial \mathscr{L}}{\partial L_2^i} = F'(L_2^i) - w_2^i - n\frac{w_m - w_2^i}{\psi} \leq 0$$

$$L_2^i \geq 0 \qquad L_2^i \frac{\partial \mathscr{L}}{\partial L_2^i} = 0$$

$$\Rightarrow \frac{\partial \mathscr{L}}{\partial L_2^i} = 0$$

$$F'(L_2^i) = w_2^i \left(1 - \frac{n}{\psi}\right) + \frac{n}{\psi} w_m \tag{2.49}$$

$$\frac{\partial \mathscr{L}}{\partial w_2^i} = -L_2^i + \frac{n}{\psi} L_2^i + \lambda_1 - \lambda_2 \leq 0 \tag{2.50}$$

$$\frac{\partial \mathscr{L}}{\partial \lambda_1} = w_2^i - \underline{w} \geq 0 \tag{2.51}$$

$$\frac{\partial \mathscr{L}}{\partial \lambda_2} = w_m - w_2^i \geq 0 \tag{2.52}$$

$$\lambda_1 \geq 0 \qquad \lambda_2 \geq 0 \qquad \lambda_1 \frac{\partial \mathscr{L}}{\partial \lambda_1} \qquad \lambda_2 \frac{\partial \mathscr{L}}{\partial \lambda_2} = 0$$

$$\Rightarrow \lambda_1 = \lambda_2 = 0 \tag{2.53}$$

Inserting (2.53) into (2.50) and ignoring the special case where $n = \psi$ yields

$$\Rightarrow L_2^i \left(\frac{n}{\psi} - 1\right) \leq 0 \tag{2.54}$$

Equation (2.54) only holds for $n/\psi \leq 1$. In this case, the firm will opt to not comply with the minimum wage requirement. However, due to the possibility of having to pay a fine as a result, the firms' labour demand curve will be kinked and curved inward at w_m because total labour costs now amount to the wage costs plus expected penalty costs. The closer the marginal productivity of secondary labour is to the minimum wage, the fewer workers will be employed below w_m. From (2.49) it can be seen, this is the case the lower the detection probability n is, or the higher the job finding rate ψ is, because in this case, average job tenure in the second sector will be lower.

If the solution to the constrained maximisation problem given above is such that the optimal wage for the firm to pay is $w_2^i \geq e$, then there will be full employment $N = L_1 + L_2$ in the economy. In that case equation (2.49) becomes

$$F'(L_2) = w_2 \left(1 - \frac{n}{\psi}\right) + \frac{n}{\psi} w_m \tag{2.55}$$

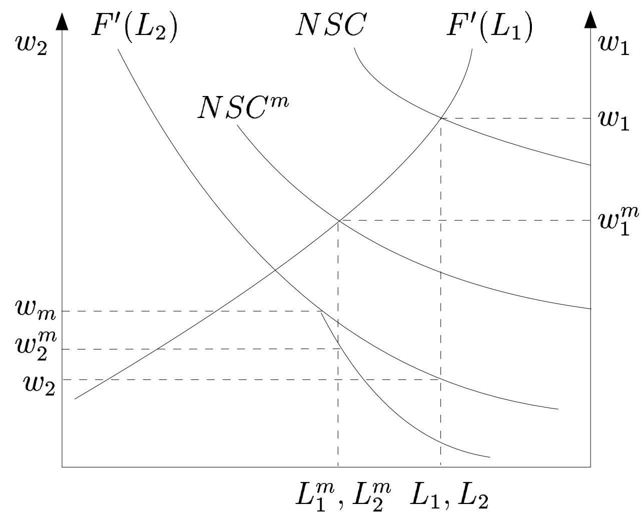
$$F'(L_1) = w_1 \tag{2.56}$$

and the no-shirking condition (2.35) changes to

$$w_1 = w_2 + \frac{e}{d}\left(\frac{qN}{N - L_1} + \rho\right) \tag{2.57}$$

Figure 2.3 shows this situation. Before there was minimum wage legislation,

Figure 2.3: Equilibrium with partial Compliance to Minimum Wage Laws



primary sector employment was determined at the intersection of the primary labour demand curve $F'(L_1)$ and the no-shirking condition, the position of which is influenced by the wage in the secondary sector. After minimum wages which are higher than the original secondary wages w_2 are introduced, the labour demand curve is kinked inwards below w_m which leads firms to reduce their secondary employment levels from L_2 to L_2^m. This reduction in secondary labour reduces the outside opportunities for someone employed in the primary sector, which means that the NSC shifts downwards to NSC^m, increasing the primary labour force size as a result. Thus in this case, the introduction of minimum wage laws not only reduces the wage differential between the two sectors, it also shifts labour from the secondary to the primary sector.

Another possibility is of course, that due to firms in the secondary sector adjusting their labour demand downwards to take the possibility of having to pay a fine into account, that secondary wages then actually fall further (below w_m). This would lead to the effect, that workers in both sectors receive lower wages, even if this situation is characterised by full employment.

Standard single sector models normally predict that minimum wage legislation normally does not improve the welfare of the targeted group, the unskilled. Firstly, if the minimum wages are higher than the currently paid wages, then the unemployment rate in this sector will increase. As a result, the increased supply of unskilled workers will seek jobs in the uncovered sector or in firms who do not comply to the law. These firms, however, will pay a wage below their previous level because they take into account the possibility of having to pay a fine. In other words, some of the unskilled previously employed now suffer a loss in wage rates due to the legislation that was designed to achieve the exact opposite. This result is offset in this dual labour market setting because the decrease in outside opportunities reduces the required no-shirking wage in the primary sector. This means that this high-paying sector can actually expand. This effect is normally not taken into account by policy makers. However, this result rests on two crucial assumptions. Firstly, the efficiency wage assumption that wage differentials and outside employment opportunities are the sole factors which determine the primary wage rate and thus employment levels in the primary sector. This implies that larger secondary sector unemployment is a concomitant of higher primary employment because it reduces primary wages. Secondly, that workers are homogeneous and can simply switch between sectors. Both of these assumptions are extreme and thus the results obtained here should be regarded with caution.

One further labour market imperfection which is often used to explain the high unemployment levels in Europe are market frictions. A whole class of matching models have been developed which analyse these frictions in more detail. These search theoretical considerations are the subject of the next chapter.

2.4 Dual Labour Markets and Matching Theory

In traditional macroeconomics, it is assumed that growth and business cycles affect all firms and sectors identically. This implies that in a recession there is a large drop in job creation accompanied by a similar increase in job destruction and further, that there is a significant increase in the flow of workers from employment to unemployment and a large decrease in the flow out of unemployment. Empirical evidence however, shows for the United States that although recessions are in fact characterised by a large increase in job destruction, job creation falls only moderately (see Davis, Haltiwanger (1990),(1992)). In Europe gross job flows actually move together and are much larger than one would expect (roughly 10% of the total labour force) given the European reputation of Eurosclerosis (see Burda, Wyplosz (1994)). Matching theories of unemployment presented here, regard the labour market as a continuous search process in which labour market institutions represent a matching technology which brings together vacancies with the unemployed.

Even if there is a large amount of job creation, with heterogeneous workers and firms, it takes time for a firm and suitable worker to be "matched" for two reasons. Firstly, due to labour market frictions, an unemployed worker will not immediately be aware of new job opportunities, and secondly, his qualifications may not be those required for the job so that firms incur costs in evaluating applicants.

So how is Europe's poor reputation regarding labour market flexibility and the resulting high unemployment compatible with the large employment flows empirically observed? Within the dual labour market approach it is possible to reconcile these two apparently opposing facts. With a two-tier labour market it is likely that the high turnover is concentrated in the secondary, low-wage sector, whilst there is high job security in the primary sector and correspondingly low job turnover. This is exactly the setup developed in SAINT-PAUL (1996a) who concentrates on frictions caused by firing costs.

Up to now, all models have assumed homogeneous workers. Although this is of course a simplification of reality, it does not alter the main general result, namely that wage differences can persist even when workers are identical. In this section, however, it is assumed that a proportion z of workers are of type 1 or skilled, and $1-z$ are of type 2 and unskilled. Thus, skilled labour here is comparable to primary labour in the previous models. When firms want to hire new workers, they first have to create a vacancy. It is assumed that these vacancies are skill-specific, i.e. that the job advertisement precisely states the type of worker required (as is realistic). In any job, a skilled worker is more productive than an unskilled worker. For this reason, firms will always prefer skilled labour. However, this makes the skilled labour market tighter, increasing the time it takes to find an appropriate worker. As it is assumed that firms can only post one vacancy per available job, this longer time duration before a type 1 (i.e skilled) vacancy is filled, reduces its value and increases the value of a type 2 vacancy. In equilibrium, the values of the two vacancy types must be identical.

The skill-specific matching function m exhibits constant returns to scale and has the number of unemployed U_j and vacancies Υ_j as its inputs, with $j = 1, 2$ denoting skilled and unskilled workers respectively. This means that the probability that a firm finds a suitable worker, θ_j is given as

$$\theta_j = \frac{m(U_j, \Upsilon_j)}{\Upsilon_j} \tag{2.58}$$

Therefore, the Bellman equation for the value to the firm of a skilled vacancy is

$$0 = -(r + \theta_1)\Upsilon_1 + \theta_1 J_1 \tag{2.59}$$

where r is the interest rate and J_1 is the present discounted value of a job held by a type 1 worker. The corresponding equation for a type 2 worker is

$$0 = -(r + \theta_2)\Upsilon_2 + \theta_2 J_2 \tag{2.60}$$

and seeing as in equilibrium $\Upsilon_1 = \Upsilon_2 = \Upsilon$ (2.59) and (2.60) can be combined to

$$\theta_1(J_1 - \Upsilon) = \theta_2(J_2 - \Upsilon) \tag{2.61}$$

Seeing as skilled workers are relatively more productive, the present discounted value of a type 1 job, J_1, is larger than the corresponding value J_2 for unskilled workers. Hence, condition (2.61) can only hold if $\theta_2 > \theta_1$, in other words, if there is a higher probability, and thus a shorter time interval, of filling a type 2 vacancy.

As in the efficiency wage literature, there is an exogenous job separation rate q. For simplicity, it is assumed that skilled workers produce an output of 2 and unskilled workers $2\beta, \beta < 1$ units and that this output is split equally between firms and workers.[9] Thus, the value of a job held by a skilled worker is

$$0 = 1 - (r+q)J_1 + q\Upsilon \tag{2.62}$$

and for a type 2 worker

$$0 = \beta - (r+q)J_2 + q\Upsilon \tag{2.63}$$

using (2.59) and (2.62) yields

$$\Upsilon = \frac{\theta_1}{r(r+q+\theta_1)} = \frac{\beta\theta_2}{r(r+q+\theta_2)} \tag{2.64}$$

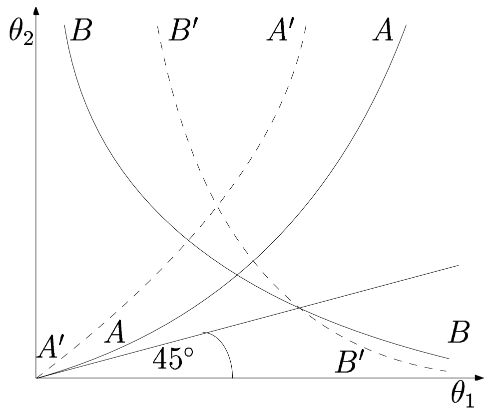
$$\Rightarrow \theta_2 = \frac{\theta_1}{\beta - \frac{\theta_1(1-\beta)}{r+q}}, \qquad \frac{\partial\theta_2}{\partial\theta_1} > 1 \tag{2.65}$$

Equation (2.65) defines an upward-sloping convex curve in the θ_1, θ_2 plane, shown as AA in Figure 2.4. The AA-curve represents the arbitrage condition for vacancies. As the slope of the curve is always greater than one, θ_2 is always larger than θ_1, reflecting the fact that in equilibrium, it must take less time to fill a type 2 vacancy. The curve also has a vertical asymptote at

$$\theta_1 = \theta_1^* = \frac{\beta(r+q)}{1-\beta}$$

[9] SAINT-PAUL (1996a) shows that the qualitative results do not change if Nash-Bargaining is assumed.

Figure 2.4: Equilibrium Job-Finding Probabilities



In this case, even with an infinitely high θ_2, i.e. a type 2 vacancy can be filled instantaneously, the value of hiring an unskilled worker will always be smaller than the value of a skilled vacancy, which means that the unskilled unemployment rate approaches one.

In a steady-state equilibrium, the number of workers leaving each sector must be equal to the number entering the sector each period

$$qL_j = \theta_j \Upsilon_j \tag{2.66}$$

where L_j is the size of the workforce of type j. Normalising the total workforce to unity and assuming that a constant fraction z are skilled means that the respective unemployment rates can be expressed as

$$U_1 + L_1 = z \tag{2.67}$$

$$U_2 + L_2 = 1 - z \tag{2.68}$$

Further, rewriting (2.58) yields

$$\theta_j = \frac{m(U_j, \Upsilon_j)}{\Upsilon_j} = m\left(\frac{U_j}{\Upsilon_j}, 1\right)$$

$$\frac{U_j}{\Upsilon_j} = g(\theta_j)$$

$$U_j = \Upsilon_j g(\theta_j) \tag{2.69}$$

where $g(\theta_j) = m^{-1}(1, \theta_j)$. As the matching function is characterised by decreasing returns in U_j and Υ_j, it must hold that given vacancies, the unemployment rate must rise disproportionately if the probability of finding a suitable worker, θ_j, is to increase. Therefore, equation (2.69) is an increasing, convex function.

Using (2.66) – (2.69) it is possible to determine the equilibrium unemployment and vacancy rates, υ_j and ν_j respectively

$$\upsilon_j = \frac{g(\theta_j)}{\theta_j/q + g(\theta_j)} \tag{2.70}$$

$$\nu_j = \frac{1}{\theta_j/q + g(\theta_j)} \tag{2.71}$$

where υ_j is an increasing and ν_j a decreasing function of θ_j. This means that the unemployment rate for unskilled workers is higher and their vacancy rate is lower than the corresponding rates for skilled workers.

The total number of jobs available in the economy, K, must be equal to the total number of jobs in the two sectors

$$K = z(1 - \upsilon_1 + \nu_1) + (1 - z)(1 - \upsilon_2 + \nu_2)$$

Inserting (2.70) and (2.71) makes it possible to derive a flow equilibrium locus, shown as BB in Figure 2.4 with

$$K = z\left(\frac{1 + \theta_1/q}{\theta_1/q + g(\theta_1)}\right) + (1 - z)\left(\frac{1 + \theta_2/q}{\theta_2/q + g(\theta_2)}\right) \tag{2.72}$$

Equation (2.72) is a downward-sloping curve with two asymptotes $\bar{\theta}_j$ at

$$K = z\left(\frac{1 + \bar{\theta}_1/q}{\bar{\theta}_1/q + g(\bar{\theta}_1)}\right) \quad \text{and } K = (1 - z)\left(\frac{1 + \bar{\theta}_2/q}{\bar{\theta}_2/q + g(\bar{\theta}_2)}\right)$$

In Figure 2.4, $\bar{\theta}_1 < \theta_1^*$ so that there is a unique equilibrium with $\theta_2 > \theta_1$. For the case that $\bar{\theta}_1 > \theta_1^*$, the arbitrage condition AA does not cross the flow equilibrium BB and all unskilled workers are unemployed.

There are two effects which are often stated in the literature as a cause of the rise particularly in the low-skilled unemployment rate. First, skill-biased technological change, which leads to an increase in demand for skilled workers, and secondly, the change in the composition of the workforce, as an ever increasing proportion acquire higher skill levels. The effects of skill-biased technological change are expressed in the context of the model here, by a decrease in β, the relative productivity of unskilled workers.

From equation (2.65) it can be seen that a lower β lowers the value of a type 2 vacancy. Therefore, in order to fulfil the arbitrage condition, the probability of finding an unskilled worker, θ_2, must rise. This is shown by the new $A'A'$ locus in Figure 2.4 which has shifted upwards relative to the original curve. From (2.70) it is clear that the unskilled unemployment rate must also rise as a result. Total unemployment is given by

$$\upsilon = z\upsilon_1 + (1-z)\upsilon_2 \tag{2.73}$$

How total unemployment changes can be seen by inserting (2.70) into (2.73) and totally differentiating

$$d\upsilon = z\mu(\theta_1)d\theta_1 + (1-z)\mu(\theta_2)d\theta_2 \tag{2.74}$$

with μ defined as

$$\mu = \frac{(\theta g'(\theta) - g(\theta))/q}{[\theta/q + g(\theta)]^2}$$

Totally differentiating the flow equilibrium condition (2.72) yields

$$dK = 0 = z\varpi(\theta_1)d\theta_1 + (1-z)\varpi(\theta_2)d\theta_2 \tag{2.75}$$

with ϖ defined as

$$\varpi = \frac{[g(\theta) - \theta g'(\theta)]/q - [g'(\theta) + 1/q]}{[\theta/q + g(\theta)]^2}$$

Combining (2.74) with (2.75) gives

$$d\upsilon = z\mu\left(1 - \frac{\mu(\theta_2)\varpi(\theta_1)}{\mu(\theta_1)\varpi(\theta_2)}\right) \tag{2.76}$$

As $g(\theta)$ is convex, it holds that $\theta g'(\theta) > g(\theta)$. As by (2.65) $\partial\theta_1/\partial\beta > 0$, then

$$\partial\upsilon/\partial\beta < 0 \quad \text{iff} \quad \frac{\theta_2 g'(\theta_2) - g(\theta_2)}{qg'(\theta_2) + 1} > \frac{\theta_1 g'(\theta_1) - g(\theta_1)}{qg'(\theta_1) + 1} \tag{2.77}$$

Seeing as both the l.h.s. and the r.h.s. are increasing functions of θ_j, and $\theta_2 > \theta_1$, condition (2.77) is always fulfilled and overall unemployment rises as the skilled workforce becomes relatively more productive.

The reason for this overall increase in unemployment lies in the nature of the matching function and in the fact that the total number of jobs is fixed in the economy. This implies that a decrease in unskilled employment leads to an increase in skilled employment. In a steady-state, a higher skilled employment rate must be accompanied by a corresponding increase in type 1 and decrease in type 2 vacancies. However, since there are decreasing returns to posting a vacancy and the skilled unemployment rate is lower to begin with, the increase in the number of matches by posting an additional skilled vacancy will be lower than the decrease in the number of matches due to the fewer type 2 vacancies. This means that the total number of matches decreases and unemployment rises.

How does the unemployment rate change if more people become skilled, i.e. z increases. Equation (2.72) is independent of z when $\theta_1 = \theta_2$. However, as both terms on the l.h.s. are decreasing in θ, the second term must be lower than the first for the relevant case $\theta_2 > \theta_1$. Thus, an increase in z raises the value of the l.h.s. of (2.72). However, as the number of jobs in the economy K, is fixed, equilibrium can only be reached if θ_2 (given θ_1) increases. The reverse is true for the case when $\theta_2 < \theta_1$. This means, that the new flow equilibrium locus $B'B'$ in Figure 2.4 rotates clockwise around the point $\theta_1 = \theta_2$. Thus, the unemployment rate will be higher for both skilled and unskilled workers for two reasons. Firstly, since there are more skilled workers, the individual probability of finding a skilled job reduces. Secondly, since firms are more likely to find a skilled worker, they are less inclined to hire an unskilled worker unless the time it takes to find one reduces drastically, which can only occur if there is a large increase in the unskilled unemployment rate. There is, however, a counteracting positive composition effect. Seeing as a higher percentage of the workforce is skilled, a higher proportion will be in the low unemployment group. Equation (2.72) is fulfilled for a unique value of θ for $z = 0$ as well as $z = 1$. Hence, the aggregate unemployment rate must be the same for these two extreme values. As z increases from zero to one, the unemployment rate must first rise and then fall back to its initial value. This is again due to the matching function. When z is small, there will only be a few type 1 vacancies as it takes a long time to fill these.

Therefore, for small z, a small increase in the number of skilled vacancies will lead to a large increase in the number of matches and by (2.58), a correspondingly large increase in the probability of finding a skilled worker, θ_1. In equilibrium, this effect must be compensated by a similar increase in θ_2 which means that both skilled and unskilled unemployment rates rise. Further, seeing as most of the workforce is unskilled, this rise in unskilled unemployment will affect most workers. The opposite holds when z is large. Summing up then, unemployment is highest when the workforce is very heterogeneous, i.e. for intermediate rather than extreme values of z. If one accepts that the share of the population that is skilled was initially low and has increased steadily over the last few decades, then part of the increased unemployment can be attributed to the increase in the supply of skilled labour. Undoubtedly, it has been unskilled unemployment though which has risen most. Within the context of the model here, this trend will continue as the share of skilled workers further increases.

The conclusions reached so far have all been based on the assumption that unskilled workers cannot be replaced by skilled ones. This assumption is relaxed now. However, firing costs D, occur if an unskilled worker is dismissed. This affects the option value of a vacancy, i.e. the value of leaving a vacancy unfilled and waiting for a more skilled applicant. Such firing costs change the value to the firm of a job held by an unskilled worker. This means that (2.63) changes to

$$0 = \beta - (r+q)J_2 + q\Upsilon + \max[0, \theta_1(J_1 - D - J_2)] \tag{2.63$'$}$$

where the last term reflects the value of posting a type 1 vacancy for a job currently held by an unskilled, type 2 worker. If this last term is positive, then from (2.59) and (2.62) one obtains

$$J_1 = \frac{r+\theta_1}{r(r+q+\theta_1)} \tag{2.78}$$

and from (2.60) and (2.63$'$)

$$\Upsilon = \frac{\theta_2\beta + \theta_1\theta_2(J_1 - D)}{(r+\theta_2)(r+\theta_1) + rq} \tag{2.79}$$

$$J_2 = \frac{\beta + q\Upsilon + \theta_1(J_1 - D)}{r+q+\theta_1} \tag{2.80}$$

Equations (2.64),(2.78) and (2.79) define the new arbitrage condition

$$\theta_2 = \frac{\theta_1}{\beta - D\theta_1} \tag{2.81}$$

which implies that firms will only replace unskilled workers if

$$D < \frac{1-\beta}{r+q} \tag{2.82}$$

i.e. the capital gain due to the higher worker productivity, denoted by the r.h.s. of (2.82), must be larger than the firing costs. The capital gain increases if the productivity difference increases (β decreases), or either the interest or quit rate fall as both of these rates reduce the discount factor of future earnings.

Inserting (2.82) into (2.81) shows that the new arbitrage condition will be flatter than the original, i.e. firms are more willing to hire unskilled workers given the probability of finding a skilled worker. If there were no firing costs the arbitrage condition would be linear, greatly reducing unskilled unemployment.

This last result of course assumes that skilled and unskilled workers are substitutes in production. Further, all the effects derived here have been for the case that the number of jobs in the economy is fixed. Both of these assumptions seem unrealistic and may have important implications for the comparative static analysis. The model should also be extended by endogenising the decision whether to become skilled or not. However and most importantly, for the first time the unemployment rate for the skilled, i.e. the primary labour force, is lower than the corresponding rate for secondary workers. This is clearly in line with all empirical evidence. It has been shown that some of the increased unemployment, for unskilled as well as the skilled, can be attributed to the higher average skill level of the workforce. This has important implications for the nature of government financed training programs. Those which purely increase the fraction of skilled workers may well be counter-effective. Only those which raise the productivity of the whole workforce, e.g. better schooling, will reduce unemployment.

In the models discussed so far, only labour market imperfections were considered. Although this is of course one of the primary reasons for unemployment, goods market imperfections can also have labour market implications. This is outlined more precisely in the next chapter.

Chapter 3

Dual Labour Markets and Uncertain Product Demand

In Chapter 2.1 the "no-shirking" efficiency wage concept was explained. As noted there, one of the main criticisms of the concept is that it does not adequately explain how firms will elicit effort from their workforce when it is clear that a worker will retire at the end of the period. In that case, dismissal no longer leads to an income loss because he would have stopped working even if he had not been dismissed. In order to overcome this weakness, the standard efficiency wage model is now extended to incorporate product demand uncertainty. Now it is not an individual worker who retires at the end of the period, but firms which may be forced to dismiss workers when demand is very low. If workers know this, they have no incentive to provide effort (as this reduces their utility) and will shirk. For this reason, the firm will want to reduce the layoff probability to a level as low as possible, as the longer the expected job tenure is, the lower will be the required wage premium to prevent workers from shirking. In other words, hiring primary workers creates linear adjustment costs in the size of the workforce as a larger workforce increases the probability that there will be layoffs as soon as product demand declines. Secondary workers on the other hand, are paid a wage that equals their marginal product and can be hired and fired at no extra cost to the firm.

To simplify the analysis without altering the qualitative results, Rebitzer, Taylor (1991) develop a model, in which demand can only take on two forms: either high or low. Firms hire homogeneous workers under two types of employment contracts, either as primary or as secondary workers. Primary workers receive a higher ("no-shirking") wage w_1 than secondary workers and also have a lower probability of being dismissed when product demand is low. Secondary workers enjoy no job security and are paid the spot market clearing wage w_2. For simplification purposes it is assumed that there are only two effort levels which workers can exert: e_h denotes high and e_l corresponds to low effort level with $e_h > e_l$. It is always possible to instantaneously detect workers who provide less than the minimal low effort. However, the probability of determining whether a worker is providing high effort or not, corresponds to the probability of detecting a worker shirking, denoted by d. As is custom in efficiency wage theory, providing effort reduces net utility of workers

$$u(w_j, e) = w_j - e, \quad j = 1, 2, \quad e \in \{e_h, e_l\} \tag{3.1}$$

For this reason, secondary workers will only exert the minimal required effort level e_l whereas higher wages induce primary workers to also exert the high effort level as long as the costs of being dismissed outweigh the benefits in net utility from receiving a high wage but having to provide the high effort level.

Uncertain product demand means that non-shirking primary workers only face the probability $\lambda < 1$ of being retained in the next period. As a flexible secondary wage means that the secondary labour market always clears. The expected discounted flow of utility for an infinitely lived non-shirking worker V_1^N, is

$$V_1^N = u(w_1, e_h) + \frac{\lambda V_1^N}{1+\rho} + \frac{(1-\lambda)V_2}{1+\rho} \tag{3.2}$$

where V_2 is the expected present value of employment in the secondary sector. A worker who shirks reduces the probability of maintaining his job from λ to $\lambda(1-d)$. Therefore, the present value of his expected lifetime utility, V_1^S, can be written as

$$V_1^S = u(w_1, e_l) + \frac{\lambda(1-d)V_1^S}{1+\rho} + \frac{(1-\lambda(1-d))V_2}{1+\rho} \tag{3.3}$$

Workers in the secondary labour market receive the wage w_2, but also have a probability ψ of obtaining a job in the high-wage sector. For this reason the present value of expected utility in the secondary sector is

$$V_2 = u(w_2, e_l) + \frac{\psi V_1^N}{1+\rho} + \frac{(1-\psi)V_2}{1+\rho} \tag{3.4}$$

In a steady state equilibrium, firms will pay the lowest primary wage possible that still ensures that the expected utility from not shirking is just as high as the utility from shirking, i.e. $V_1^N = V_1^S$. The primary wage that firms will pay can be derived using equations (3.1) – (3.4) as

$$w_1 = w_2 + (e_h - e_l) + \frac{(e_h - e_l)(1 + \rho + \psi - \lambda)}{d\lambda} \tag{3.5}$$

From equation (3.5) it can be seen that the net utility of a primary worker, $w_1 - e_h$, exceeds that of a secondary worker, $w_2 - e_l$. Therefore, the equilibrium is characterised by an excess supply of workers wanting to work in the primary sector. Further, it can be seen from equation (3.5) that the primary wage is lower the higher the retention probability λ is.

The retention probability λ is endogenously determined and depends, via the product demand a firm faces in each period, on the revenue R of a firm. This is given by

$$R = pF(L_1), \qquad F' > 0, F'' < 0$$

where in each period the output price p is now a draw from a known distribution. Firms will offer $\tilde{L}_1$ primary work contracts, specifying both the required effort as well as the wage. In addition, the firm states the size of the primary workforce it will retain at the end of the period, L_1, which depends on the realisation of p. It is only after the contracts have been signed that the value of p is determined and the corresponding number of workers are retained.

In order to simplify the analysis, it is assumed that there are only two different states of p

$$p = \begin{cases} p^H > 1, & \text{with probability } \phi \\ 1, & \text{with probability } 1 - \phi \end{cases} \tag{3.6}$$

Therefore, the expected profits of a firm are

$$E(\pi) = \phi[p^H F(\tilde{L}_1) - w_1 \tilde{L}_1] + (1 - \phi)[F(L_1) - w_1 L_1] \tag{3.7}$$

If firm demand turns out to be low, then of the $\tilde{L}_1$ workers the firm hired at the beginning of the period, it will retain L_1 to maximise expected profits

$$\frac{\partial E(\pi)}{\partial L_1} = (1 - \phi)[F'(L_1) - w_1] - [\phi \tilde{L}_1 + (1 - \phi) L_1] \frac{\partial w_1}{\partial \lambda} \frac{\partial \lambda}{\partial L_1} \stackrel{!}{=} 0 \tag{3.8}$$

It is assumed that the probability of being laid off if demand is low is the same for each worker initially offered a primary contract, i.e. that the decision which workers to dismiss is random. Therefore, the retention probability can be written as

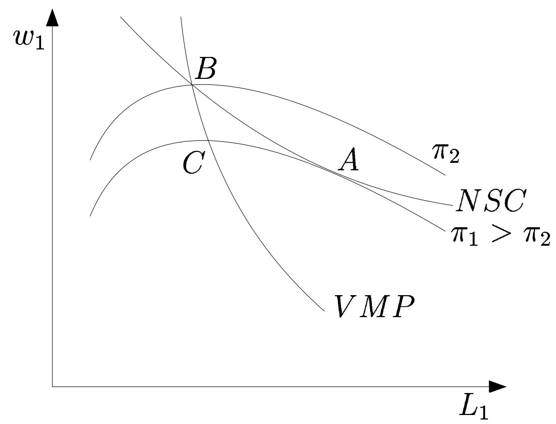
$$\lambda = \phi + (1 - \phi)\frac{L_1}{\tilde{L}_1} \tag{3.9}$$

Finally, using equation (3.9) as well as the no-shirking wage equation (3.5) the optimal number of retained workers must satisfy

$$F'(L_1) - w_1 + \frac{(e_h - e_l)(1 + \rho + \psi)}{d\lambda} = 0 \tag{3.10}$$

It is therefore optimal, i.e. less costly for firms to pay a wage above marginal productivity and to hoard primary labour if demand is low. Were the firm to fire all primary workers as soon as demand is low, then this would reduce the retention probability and lead to higher wage costs to prevent workers from shirking. The equilibrium condition is depicted graphically in Figure 3.1.

Figure 3.1: Equilibrium Wage and Employment when Product Demand is Low



The NSC curve describes the relationship between the primary no-shirking wage and the number of retained primary workers. If the no-shirking condition were not binding, firms would equate the value of marginal product VMP with the wage, for example points A or B in Figure 3.1. However, although the primary wage is higher in point C than in the equilibrium point A, the number of retained workers is so low, that point C is incompatible with the no-shirking condition. At point C the size of the workforce is relatively small. This means, that the chances of being randomly dismissed when demand is low is relatively high. In order not to shirk with this high dismissal probability, wages would need to be set at the level given by point B. Point B, however, does not represent a steady state, as the firm can increase its profits by hiring (or retaining) more workers at a lower wage.

Turning to the case when product demand is high, the firm will hire primary workers until the expected marginal benefit equals the expected marginal cost and retain all workers. From equation (3.7) this can be written as

$$\phi\left[p^H F'(\tilde{L}_1)\right] = \phi w_1 + \frac{\partial w_1}{\partial \lambda}\frac{\partial \lambda}{\partial L_1}[\phi \tilde{L}_1 + (1-\phi)L_1] \tag{3.11}$$

where the r.h.s. denotes the marginal cost of hiring an additional primary worker. This is compromised of the expected wage plus the effect an additional worker has on wages. The more workers $\tilde{L}_1$ a firm offers primary wage contracts to, the lower is the probability of being retained when demand is low. This in turn increases the required no-shirking wage. Inserting equations (3.5) and (3.9) into (3.11) yields

$$p^H F'(\tilde{L}_1) = w_2 + (e_h - e_l) + \frac{(e_h - e_l)(1+\rho+\psi-\phi)}{d\phi} \tag{3.12}$$

From equation (3.9), it can be seen that $\lambda > \phi$. Using this result to compare the primary wage equation (3.5) with the optimal number of primary workers when demand is high (3.12), shows that at the optimal level of employment the primary wage w_1 is lower than the expected value of the marginal product of a primary worker. In other words, when demand is high, the firm enjoys an extra rent on labour. However, when demand is low, firms will have labour costs which exceed the benefits of a primary worker.

Due to the fact that secondary labour can be fired at no extra (direct or indirect) costs to the firm, it may be optimal for firms to hire secondary workers in response to product market variations. In order to show this, the revenue function is modified slightly as a function of "effective" total labour $\hat{L}$

$$\hat{L} = L_1 + \beta L_2 \tag{3.13}$$

where β denotes the marginal productivity of secondary labour L_2, relative to that of a primary employee. The value of β is critical in determining the labour composition of the firm. If secondary labour productivity is so high that unit labour costs L_2/β are lower than the lowest primary wage w_1 needed to prevent shirking, then the firm will only hire secondary labour. This condition holds for

$$\frac{w_2}{\beta} < w_2 + (e_h - e_l) + \frac{(e_h - e_l)(\rho + \psi)}{d} \tag{3.14}$$

The other extreme case is when secondary labour productivity is so low that secondary labour costs always exceed the expected marginal costs of hiring primary workers. This case can be written as

$$\frac{w_2}{\beta} > w_2 + (e_h - e_l) + \frac{(e_h - e_l)(1 + \rho + \psi - \phi)}{d\phi} \tag{3.15}$$

This means that there is a corridor in which a dual structure arises within a firm, i.e. that it hires secondary as well as primary labour simultaneously

$$\begin{aligned} w_2 + (e_h - e_l) + \frac{(e_h - e_l)(\rho + \psi)}{d} &< \frac{w_2}{\beta} \\ &< w_2 + (e_h - e_l) + \frac{(e_h - e_l)(1 + \rho + \psi - \phi)}{d\phi} \end{aligned} \tag{3.16}$$

This dual structure arises when secondary labour costs are greater than the no-shirking wage when $\lambda < 1$. There are two opposing effects which the firm needs to consider. On the one hand, primary workers are more productive and thus cheaper than secondary labour. On the other hand, as can be seen from equation (3.9), an increase in the number of workers hired ex-ante $\tilde{L}_1$, lowers the retention probability and inversely, via equation (3.5), raises the primary wage rate. This means, that an increase in the number of primary workers increases the marginal labour costs. In this case the firm will want to hire as many workers as are needed when demand is low. As soon as demand is high though, it will be optimal for the firm to hire secondary workers because these are cheaper than primary workers when the firm uses layoffs to adjust to demand.

The above setup is modified in SAINT-PAUL (1996b) where product demand can take on any value. It is modelled as a series of temporary stochastic shocks, with the size of these shocks determining the employment in each period. As above, up to a certain critical value, it is less costly for firms to pay their workers a wage above their marginal productivity rather than firing them but having to increase the wage rate as a result. However, if demand is below this critical value, the firm will be forced to reduce the size of its workforce. Similarly, there will exist a certain upper ceiling for demand, above which it will be optimal for the firm to hire more labour. Although there is strictly speaking no unemployment in the secondary sector, as the wage here is determined on the spot market, it is always secondary workers who are fired first so that is these who suffer most from unemployment.

As can be seen then and similar to the class of efficiency wage models, firms have an incentive to deviate from their short-term profit maximising labour force size. It also follows that primary and secondary sector jobs follow different wage patterns. Whereas the wage of the former is only linked to the worker's marginal productivity in the long-run, secondary wages are closely related and sensitive to productivity changes. Further, if demand fluctuations become too large, then firms will have an incentive to reduce their primary labour size and increase their demand for secondary workers.

As stated at the very outset, dual labour markets are characterised by two wage levels in equilibrium. However, in standard neoclassical economics, as soon as open economies are analysed, not only is there only one wage in each country, but in a Heckscher-Ohlin framework wages in all countries are identical as well. The final chapter presents a dual labour market model in which even with international trade, wage differentials persist.

Chapter 4

Dual Labour Markets in Open Economies

Until now we have assumed a closed economy. This assumption is relaxed now. Following COPELAND (1989), the efficiency wage concept is integrated into a Ricardian model of international trade.[1]

The setup is very similar to that of BULOW, SUMMERS (1986) presented in Chapter 2.1. The main difference between the two is that here both sectors of the economy employ primary and secondary labour, or in other words, there are primary and secondary jobs available in both sectors. That is, the production functions for the two goods are

$$X_1 = F^1(L_1^1, L_2^1) \tag{4.1}$$

$$X_2 = F^2(L_1^2, L_2^2) \tag{4.2}$$

where X_1 and X_2 are the total amounts produced in the respective sectors, and L_i^j are the number of jobs of type j, i.e. primary or secondary, working in sector i. The production functions are assumed to be linearly homogeneous but the primary sector relatively intensively uses primary labour as its input.

Workers obtain utility from consuming either of the two goods but providing

[1] See also, for example, BHAGWATI, SRINIVASAN (1971),KHAN (1979),SCHWEINBERGER (1998) for three very different approaches to dual labour markets in open economies.

effort reduces their utility according to the functional form

$$u = u(x_1, x_2)h(e) \tag{4.3}$$

where $h(e)$ is the effort function. Workers can either shirk or provide effort. If they do shirk, the utility gain is α so that

$$h(0) = 1 + \alpha, \qquad \alpha > 0$$
$$h(1) = 1$$

from which the indirect utility function can be written as

$$\Theta(p_1, p_2, w_j, e) = \frac{w_j h(e)}{\tilde{p}} \tag{4.4}$$

Here, Θ denotes indirect utility, p_i are the prices in the two sectors, w_j is the wage rate for type j jobs and $\tilde{p}$ is the macroeconomic price index.

As in the standard efficiency wage model, the market for secondary jobs is perfectly competitive. All primary sector jobs have an exogenous job separation rate q and due to imperfect monitoring, shirkers are only detected with the probability d. This means that if agents' subjective discount rate is given by ρ and they have infinite lives, the expected present value of lifetime utility for a non-shirker V_1^N is

$$\rho V_1^N = \frac{w_1}{\tilde{p}} - q(V_1^N - V_2) \tag{4.5}$$

and similarly V_1^S

$$\rho V_1^S = \frac{w_1(1+\alpha)}{\tilde{p}} - (q+d)(V_1^S - V_2) \tag{4.6}$$

for a shirker. Monitoring is perfect for secondary jobs, so that workers here never shirk. The number of new primary jobs opening in each period is $(q+d)L_1$ and the probability of obtaining such a job is $(q+d)\Lambda$ where Λ is the ratio of primary to secondary jobs. Therefore, the Bellman equation for a worker currently employed in a secondary job is

$$\rho V_2 = \frac{w_2}{\tilde{p}} - (q+d)\Lambda(V_1^N - V_2) \tag{4.7}$$

In equilibrium, in order for workers not to shirk, the costs from shirking, i.e. the higher probability of losing a primary job must be at least as large as the utility gained from shirking. Seeing as no employer will pay a higher wage than necessary to prevent shirking, in equilibrium it will hold that

$$V_1^N = V_1^S \tag{4.8}$$

Using equations (4.5) – (4.8) it is possible to derive the wage differential between the primary and secondary wage as

$$w_1 = \frac{dw_2}{d - \alpha(\rho + q + \Lambda(q + d))} \tag{4.9}$$

The home country is small and thus cannot influence the world prices p_1 and p_2. From the production functions (4.1) and (4.2) it is possible to derive the cost function. Applying Shephard's Lemma then yields the factor demand functions

$$k_1^1(w_1, w_2)X_1 + k_1^2(w_1, w_2)X_2 = L_1 \tag{4.10}$$

$$k_2^1(w_1, w_2)X_1 + k_2^2(w_1, w_2)X_2 = L_2 \tag{4.11}$$

$$N \geq L_1 + L_2 \tag{4.12}$$

where N is the size of the total workforce and k_j^i the derivative of the cost function in sector i with respect to labour type j. This means that k_j^i can be interpreted as the input coefficients.

In order to derive the labour force shares in both sectors it is first necessary to use (4.9) to establish a relationship between the relative wage $\omega = w_1/w_2$ and the relative labour shares Λ

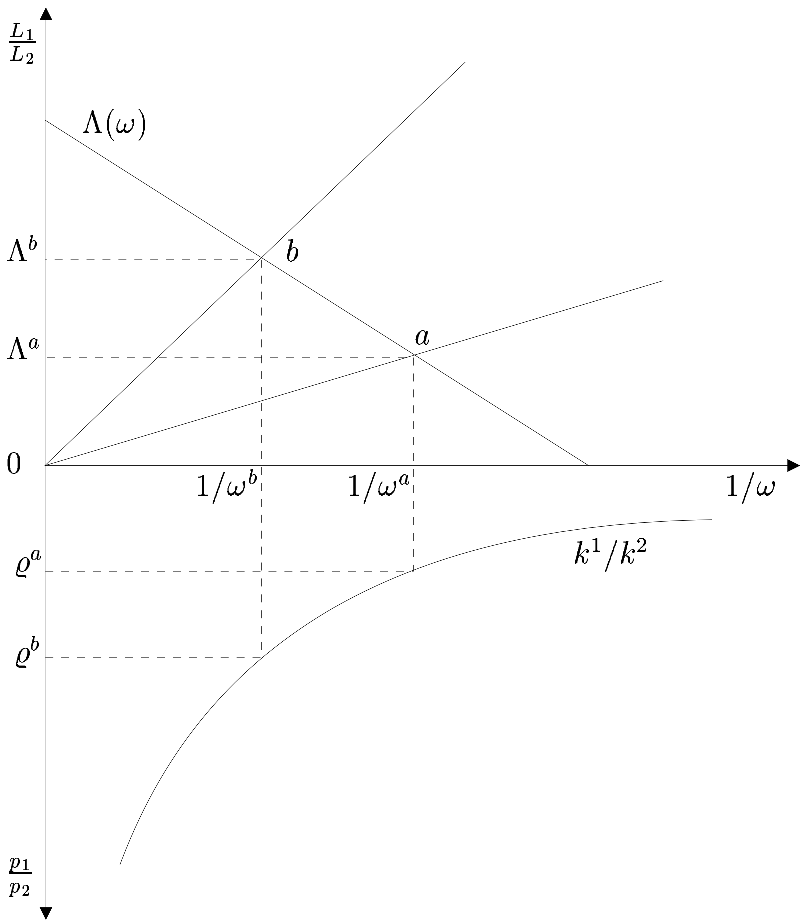
$$\Lambda = \frac{d}{\alpha(q + d)}\left(1 - \frac{1}{\omega}\right) - \frac{\rho + q}{q + d} \equiv \Lambda(\omega) \tag{4.13}$$

The relative labour shares in the two sectors can be determined from above as the ratio of the input coefficients

$$\Lambda_1 = \frac{k_1^1(\omega)}{k_2^1(\omega)} \tag{4.14}$$

$$\Lambda_2 = \frac{k_1^2(\omega)}{k_2^2(\omega)} \tag{4.15}$$

Figure 4.1: Determination of the Relative Number of Workers as a Function of Relative World Prices



Equations (4.13) – (4.15) are shown in Figure 4.1 from which it is possible to derive the price range for which the home economy will produce both goods. The top half of Figure 4.1 determines the equilibrium labour shares. From the bottom half, where ϱ denotes world relative prices it is possible to determine which goods are produced. At point a and at any price level below ϱ^a, only the secondary good X_2 will be produced. Similarly, any price level higher than that given by ϱ^b means that the economy only produces the primary good. For any price range between ϱ^a and ϱ^b the economy produces both goods. In this range it holds that $\Lambda_1 \geq \Lambda \geq \Lambda_2$ which is consistent with the fact that the primary sector is relatively intensive in type 1 jobs.

For the case that both goods are produced in the economy, the above model is similar to the Heckscher-Ohlin model. Therefore, it is also possible to apply the Stolper-Samuelson theorem. An increase in p_1, the price of the good which intensively uses primary labour, means that more of this good will be produced, hence increasing relative demand for primary jobs and raising the wage they receive. At the same time, the real wage rate for secondary labour must decline.

As in the original BULOW, SUMMERS (1986) analysis, a primary wage subsidy would increase welfare as the subsidy can eliminate the distortion caused by the higher efficiency wages being paid for primary jobs. Although this would be a first-best intervention, welfare enhancing effects are also possible by imposing small tariffs. For example, if a tariff is imposed on the primary good, then from above, this increases the primary wage and therefore the utility of primary workers. However, as can be seen from (4.7), a higher primary wage also indirectly increases welfare of secondary workers, as these have a positive probability of entering the primary sector during their lives. Even if the tariff creates a consumption distortion effect, this is initially offset by the increase in government revenue.

The above policy recommendations are of course based on a model with an extremely simple production technology and perfectly competitive goods markets. All of these assumptions need to be given up, before more exact welfare effects can be obtained. Further, the critique which applies to efficiency wage models in general, naturally also applies here. However, even if following up these extensions would add richness and precision to the model, it seems unlikely that the central claims would change as a result.

Chapter 5

Conclusion

As shown in this survey, all modern unemployment theories can be integrated into a dual labour market setting. For this reason, there are only few characteristics which apply to all dual labour market models and the criticism that applies to, for example, efficiency wage theory, also applies to dual labour market models based on efficiency wages. This also means that all empirical tests of the theory strictly depend on the underlying model being tested. Numerous work by DICKENS,LANG (1993,1988) amongst many other contributions and by KRUEGER, SUMMERS (1988) mainly test the efficiency wage variants of dual labour markets and find considerable evidence proving the existence of such a segmented structure. Similarly, GIBBONS, KATZ (1992) have also found evidence in support of the dual labour markets when applied to a matching model. KAHN (1998) not only tests the theory in the context of collective bargaining but also provides evidence from US and European data emphasizing the validity of the theory, although the inter-industry wage differences are smaller in Europe than in the US. All these findings clearly illustrate that the standard neoclassical economics explanation by which wage differences between sectors are solely due to different working conditions or worker heterogeneity cannot explain all of the empirical evidence and thus, that dual labour market theory is a necessary and viable extension to modern labour economics.

Whereas the earliest dual labour market models differed sharply from mainstream economics, a fact which led to the repudiation of the theory at first, the differences between standard and dual labour market economics have become less and less distinctive which, as Dickens, Lang (1988) state, has led to the reemergence of the theory. Recent dual labour market models all have as their base some form of labour market imperfection and in this respect there is no contradiction to the mainstream analysis. The two approaches differ mainly in that in standard economics different wage levels are mainly attributed to worker heterogeneity, whereas in dual labour market models emphasis is placed more on heterogeneity amongst jobs.

However, with unemployment rates in all industrialised countries differing considerably between workers employed in low-skill as opposed to high-skill jobs, it is obvious that there must be different economic factors at work in these two sectors. It is this underlying basic intuition which justifies further research on dual labour markets. This does not mean that future work in this field should not continue to use basic results derived from single sector analysis into account. For example, it seems likely that a single firm simultaneously employs low- as well as high-skilled workers but pays above average wages to both. The first fact can only be explained if heterogeneous workers are allowed for, and the second fact underlines the validity of dual labour market theory. This leads to the conclusion that neither the theory of compensating wage differentials nor segmented labour market theory *on their own* can explain all of the unemployment currently experienced especially in Europe. This makes the integration of the two an all more important goal.

References

Bhagwati, J., Srinivasan, T. (1971): Alternative Policy Rankings in a Large, Open Economy with Sector-Specific Minimum Wages. *Journal of Economic Theory* 11, 356–371.

Bulow, J. I., Summers, L. H. (1986): A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment. *Journal of Labor Economics* 4, 376–414.

Burda, M. (1992): 'Wait Unemployment' in Europe. *Economic Policy* 7, 393–425.

Burda, M., Wyplosz, C. (1994): Gross Labor Market Flows in Europe: Some Stylized Facts. *European Economic Review* 38, 1287–1325.

Cain, G. G. (1976): The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey. *Journal of Economic Literature* 14, 1215–1257.

Calvo, G. A. (1978): Urban Unemployment and Wage Determination in LDC's: Trade Unions in the Harris-Todaro Model. *International Economic Review* 19, 65–81.

Clark, K. B., Summers, L. H. (1979): Labor Market Dynamics and Unemployment: A Reconsideration. *Brookings Papers on Economic Activity* 1, 13–60.

Copeland, B. R. (1989): Efficiency Wages in a Ricardian Model of International Trade. *Journal of International Economics* 27, 221–244.

Davis, S., Haltiwanger, J. (1990): Job Creation, Job Destruction, and Job Reallocation over the Cycle. In: Blanchard, O. J., Fisher, S. (Eds.), *NBER Macroeconomics Annual 5*, Camb., MA, 123–168.

DAVIS, S., HALTIWANGER, J. (1992): Gross Job Creation, Gross Job Destruction, and Employment Reallocation. *Quarterly Journal of Economics* 107, 819–864.

DICKENS, W. T., LANG, K. (1988): The Reemergence of Segmented Labor Market Theory. *American Economic Review* 78, 129–134.

DICKENS, W. T., LANG, K. (1993): Labor Market Segmentation Theory: Reconsidering the Evidence. In: DARITY, W. (Ed.), *Controversies in Labor Economics*, 141–180.

DOERINGER, P. B., PIORE, M. J. (1971): *Internal Labor Markets and Manpower Analysis*. Lexington, MA.

GIBBONS, R., KATZ, L. (1992): Does Unmeasured Ability Explain Inter-Industry Wage Differentials. *Review of Economic Studies* 59, 515–535.

HARRIS, J. R., TODARO, M. P. (1970): Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review* 60, 126–142.

JONES, S. R. (1987a): Minimum Wage Legislation in a Dual Labor Market. *European Economic Review* 31, 1229–1246.

JONES, S. R. (1987b): Screening Unemployment in a Dual Labor Market. *Economics Letters* 25, 191–195.

KAHN, L. M. (1998): Collective Wage Bargaining and the Interindustry Wage Structure: International Evidence. *Economica* 65, 507–534.

KATZ, L. F. (1986): Efficiency Wage Theories: A Partial Evaluation. In: FISCHER, S. (Ed.), *NBER Macroeconomics Annual 1*, Camb., MA, 235–276.

KHAN, M. A. (1979): The Harris-Todaro Hypothesis and the Heckscher-Ohlin-Samuelson Trade Model. A Synthesis. *Journal of International Economics* 10, 527–547.

KRUEGER, A. B., SUMMERS, L. H. (1988): Efficiency Wages and the Inter-Industry Wage Structure. *Econometrica* 56, 259–293.

MATUSZ, S. J. (1994): International Trade Policy in a Model of Unemployment and Wage Differentials. *Canadian Journal of Economics* 27, 939–949.

McDonald, I. M., Solow, R. M. (1985): Wages and Employment in a Segmented Labor Market. *Quarterly Journal of Economics* 100, 1115–1141.

Ordover, J. A., Shapiro, C. (1984): Advances in Supervision Technology and Economic Welfare: A General Equilibrium Analysis. *Journal of Public Economics* 25, 371–389.

Rebitzer, J. B., Taylor, L. J. (1991): A Model of Dual Labor Markets When Product Demand is Uncertain. *Quarterly Journal of Economics* 106, 1373–1383.

Romer, D. (1996): *Advanced Macroeconomics*. New York et.al.

Saint-Paul, G. (1996a): Are the Unemployed Unemployable? *European Economic Review* 40, 1501–1519.

Saint-Paul, G. (1996b): *Dual Labor Markets: A Macroeconomic Perspective*. Camb., MA.

Salop, S. C. (1979): A Model of the Natural Rate of Unemployment. *American Economic Review* 69, 117–125.

Schweinberger, A. G. (1998): Dual Labour Markets, Unemployment and Trade Gains/Losses in Developing Countries. In: Koch, K.-J., Jaeger, K. (Eds.), *Trade, Growth, and Economic Policy in Open Economies. Essays in Honour of Hans-Jürgen Vosgerau*, 31–46.

Sellekaerts, B. H., Welch, S. W. (1984): An Econometric Analysis of Minimum Wage Noncompliance. *Industrial Relations* 23, 244–259.

Shapiro, C., Stiglitz, J. E. (1984): Involuntary Unemployment as a Worker Discipline Device. *American Economic Review* 74, 433–444.

Solow, R. M. (1979): Another Possible Source of Wage Rigidity. *Journal of Macroeconomics* 1, 79–82.

Solow, R. M. (1986): Unemployment: Getting the Question Right. *Economica* 53, S23–S34.

Wachter, M. L. (1974): Primary and Secondary Labor Markets: A Critique of the Dual Approach. *Brookings Papers on Economic Activity* 3, 637–680.