

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is the author's post-print version of an article published in the **Journal of Socio-Economics, 37 (2)**

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/id/eprint/77126>

Published article:

Sawyer, M and Spencer, D (2008) *On the definition of involuntary unemployment*.
Journal of Socio-Economics, 37 (2). 718 - 735. ISSN 1053-5357

<http://dx.doi.org/10.1016/j.socec.2006.12.049>

On the definition of involuntary unemployment

Malcolm Sawyer and David Spencer

Economics Division,

Leeds University Business School

March 2006

Abstract:

Keynes gave two definitions which could be seen more as tests for the existence of involuntary unemployment than definitions. It is argued that Keynes's definitions are inadequate as tests of unemployment and in many circumstances would fail to reveal the existence of unemployment. We argue that unemployment as a theoretical construct is always involuntary, and the appearance of voluntary unemployment arises only through the mismeasurement of unemployment. We argue that involuntary unemployment has to be viewed as a macroeconomic concept rather than a microeconomic one. In a collective sense, the degree of unemployment is not the outcome of decisions made by the individuals concerned.

JEL classifications: E24, J64

Keywords: Unemployment, involuntary unemployment, demand for labour

Address for correspondence:

Economics Division,

Leeds University Business School,

The University of Leeds,

Leeds LS2 9JT

UK

emails : mcs@lubs.leeds.ac.uk, das@lubs.leeds.ac.uk

On the definition of involuntary unemployment*

Malcolm Sawyer and David Spencer

Economics Division, Leeds University Business School

1. Introduction

Ever since Keynes wrote his *General Theory* (hereafter GT), there has been interest amongst economists in the concept of involuntary unemployment. Keynes defined this concept in two separate ways in his GT and debate has continued as to whether these two definitions are compatible. But there remains a wider concern over whether, in fact, Keynes defined involuntary unemployment at all (see de Vroey, 2004). Keynes sought to indicate the conditions under which involuntary unemployment would exist in the real world but did not offer a theoretical definition of involuntary unemployment as such. This raises a broader issue as to the development of macroeconomics since Keynes, in so far as little regard has been given to the question of who is to be classified as involuntarily unemployed at any given point in time.

This question provides the motivation for the following paper, which focuses directly upon the conception of involuntary unemployment. It considers the two definitions of involuntary unemployment offered by Keynes in his GT, and argues that the definitions are significantly different and are more akin to tests of the existence of unemployment than definitions. Further, it is argued that each of the definitions is model specific and would only signal the existence of involuntary unemployment under certain rather restrictive assumptions. As the definitions of Keynes are regarded as flawed, we seek to develop our own definition of unemployment. It is argued that unemployment as a concept always involves denial of the exercise of choice and should be seen as intrinsically involuntary. In addition, it is argued that involuntary unemployment is a macroeconomic rather than a microeconomic concept.

The paper is organised as follows. Section 2 sets out some terminological and conceptual issues which inform the ensuing discussion. Sections 3 and 4 examine Keynes's two definitions of involuntary unemployment. Section 5 addresses the issue of imperfect competition as a factor in the explanation of unemployment. Section 6 presents a new definition of involuntary unemployment. Section 7 concludes.

2. Issues of terminology and of concept

A major difficulty which has plagued macroeconomics and the discussion of unemployment arises from the status and meaning of the curve which relates the real product wage and the level of employment arising from decisions made by firms. This curve, which we will label the 'employment real wage relationship' (hereafter ERWR) may refer to the firm level or be aggregated up to the industry or economy level. It is, of course, often called the demand for labour curve (or function, schedule etc.) but that terminology is, in general, misleading. The general notion of a demand function (which in the terminology of Clower (1965) would refer to 'notional demand') is that it indicates how much would be demanded at each given price and that movements along the demand function come from parametric variations in price. Yet, the employment real wage relationship does not have those features, and specifically the real (product) wage is generally set by firms as a consequence of their pricing decisions and movements along the ERWR arise from, *inter alia*, movements in the level of aggregate demand.

In the neo-classical presentation, in a situation of perfect competition, a firm is assumed to maximise profits, from which the condition of marginal product of labour equal to the real wage is derived. This is then translated into the demand for labour curve under the assumption that price and money wages are given so far as the firm is concerned, and that movements along the demand curve arise from variations in the price:wage ratio. Further, there is the implicit assumption that individual firms are able to sell the resulting output and

that the only variables influencing demand for output are relative prices (and not the level of aggregate demand).

The perspective of Keynes (1936) of relevance here involves two elements. First, the level of employment is determined by the level of effective demand. Second, firms adhere to the rule of price equals marginal cost (though a constant mark-up of price over marginal cost would not change matters significantly, provided that firms still operate on the rising portion of their marginal cost curve). Keynes (1936, p.294) argued that ‘in a single industry its particular price-level depends partly on the rate of remuneration of the factors of production which enter into its marginal cost, and partly on the scale of output. There is no reason to modify this conclusion when we pass to industry as a whole’. Keynes (*ibid.*, pp.302-3) continued that ‘The cost-unit, or, subject to the above approximation, the wage-unit, can thus be regarded as the essential standard of value; and the price-level, given the state of technique and equipment, will depend partly on the cost-unit and partly on the scale of output, increasing, where output increases, *more* than in proportion to any increase in the cost unit, in accordance with the principle of diminishing returns in the short period’. It can be seen from this that what Keynes terms as the ‘first classical postulate’, that the real product wage equals the marginal product of labour, will be met. Yet, in the GT, ‘the marginal product curve is not the demand curve for labor’ (to use the first part of the title of Davidson’s (1983) article). The real wage is set by firms from their pricing decisions and changes in the real wage, employment combination would arise from changes in the level of effective demand. That is, a change in effective demand leads to a change in employment, which in turn causes a change in the real wage: here the direct causation runs from employment to real wages, not the other way around, as is assumed in neoclassical theory. Keynes believed that firms would operate on their downward-sloping marginal product of labour curve. Writing to Ohlin in 1937, Keynes (1973, p.190) asserted that ‘I have always regarded decreasing physical returns in the

short period as one of the few incontrovertible propositions of our miserable subject'. Hence increased use of labour would be associated with a lower marginal product and thereby real wage, though Keynes viewed any increase in the use of labour as arising from an increase in demand.

In the case of imperfect competition, the nature of the relationship between employment and the real wage is even clearer (for further elaboration see Sawyer, 1992, 2002). Take the simplest case, that of monopoly. The short-run objective function of the enterprise is given as follows $\Pi = P(q, X).q - W.l$ where $q = f(l, k)$: P is price charged by the firm, q enterprise output, X is a vector of variables influencing the demand facing an enterprise including the level of aggregate demand, W the enterprise nominal wage, l the labour input, k is the capital stock. Lower case letters refer to quantities at the enterprise level. Material inputs are omitted since their inclusion would complicate the analysis without being of importance to the points which we wish to explore here.

Using the level of employment as the key decision variable, with the capital stock and Z held constant, the first order condition for profit maximisation yields:

$$\frac{(e-1)}{e} \alpha (l)^{\alpha-1} k^{1-\alpha} f'(l, k^{1-\alpha}) = \frac{W}{P(q, Z)} \quad (1)$$

where f' is the first partial derivative of f with respect to the first argument and e is the elasticity of demand. This first order condition has often been regarded as the demand for labour schedule but it is clearly not. The enterprise sets the real product wage through its decisions over price. Further, this equation provides only a point outcome: it is an equation in l (k being treated as given at this point, P, q being functions of l) which can be solved to give the level of employment, from which the level of output, real product wage and price can be derived. In order to map out a relationship between the real product wage and employment based on equation (1) it is necessary to vary some exogenous variable, and the level of aggregate demand is the one on which we focus. Movements in the level of aggregate

demand would generate movements in employment, and in turn, the real product wage. Making such variations in Z would lead to a relationship as sketched in Figure 1, where it is assumed that the function f is such that it initially displays increasing returns to labour and then diminishing returns. If the elasticity of demand is constant, then the relationship in Figure 1 is merely the inversion of the U-shaped short run cost curve. The relationship derived from equation (1), and its aggregate counterpart, is drawn as the ERWR curve in Figure 1.

Figure 1 near here

The ERWR curve (whether under perfect or imperfect competition) differs conceptually from the notional demand for labour curve. The ERWR and the demand for labour curves may have a similar shape and there has often been confusion between them. The particularly significant difference between them is that movement along the demand for labour curve would come from variations in the real wage facing firms (under the assumption that firms could sell the resulting output) whereas movement along the ERWR comes from variations in the level of aggregate demand. This has obvious policy implications, in so far as the cause of fluctuations in employment differs between the two cases. The above discussion enables us to proceed to an examination of Keynes's two definitions of involuntary unemployment.

3. Keynes's first definition

The first definition of involuntary unemployment provided by Keynes was as follows: '*Men are involuntarily unemployed if, in the event of a small rise in the price of wage-goods relatively to the money wage, both the aggregate supply of labour willing to work for the current money-wage and the aggregate demand for it at that wage would be greater than the existing volume of employment*' (Keynes, 1936, p.15, emphasis in original). This is not a definition, only that involuntary unemployment can be said to exist, if lower real wages are associated with an increase in both the demand for and supply of labour above the existing

level of employment. But it appears to carry with it connotations for the causes of unemployment: that is, a position of involuntary unemployment is due to the real wage being too high to clear the labour market. Firstly, it can be noted here that Keynes is speaking of unemployment of people rather than underemployment in terms of hours worked. It could be said that, when he spoke of the aggregate supply of and aggregate demand for labour, this referred to the supply of and demand for numbers of people under the assumption that average hours worked are constant. Alternatively, it could be interpreted in terms of hours worked with the distribution between numbers and hours to be determined. Secondly, the emphasis is very much on adjustment in the real wage, with a reduction in the real wage, brought about by higher prices, being associated with a rise in both the quantities demanded and supplied of labour. This would seem to suggest that the rise in employment brought about by an increase in demand is accompanied by a fall in the real wage.

As Darity and Young (1997, pp.22-3) argue, Keynes's first definition of involuntary unemployment 'is identical with the straightforward notion of an excess supply of labour under the following narrow conditions: 1) the aggregate demand schedule for labour is negatively related to the real wage; 2) the aggregate supply schedule for labour is positively related to the real wage; 3) the only endogenous variable that both schedules depend upon is the real wage rate; 4) there is only one sector in the macroeconomy'. This quote is suggestive of a competitive demand and supply framework with the real wage above the market clearing level (and the short-side of the labour market determining employment). A lower real wage then is associated with a greater quantity of labour demand than initially (where labour demand equals actual employment) and to a fall in the quantity of labour supplied (assuming the labour supply curve is upward sloping), but with the available supply of labour remaining greater than initial employment. Full employment is restored in this case once the labour market clears.

Joan Robinson had much earlier pointed out that ‘Mr. Keynes’ definition of involuntary unemployment is not strictly accurate [it] is appropriate only to short-period conditions, when a fall in real wages is a symptom of an increase in effective demand, and it is formally correct only for cases in which the elasticity of supply of labour in terms of real wages is zero or positive’ (Robinson, 1937, p. 172, fn. 1). The restrictiveness of this definition, as Robinson went on to show, can be readily seen by relaxing these assumptions (see Ambrosi, 2003, pp.91-2). Thus, the implied equality between labour market clearing and the absence of involuntary unemployment given by Keynes in his first definition can be shown to be invalid under (not unrealistic) conditions relating to the slopes of the ‘demand’ and supply of labour curves, as will be shown below.

It may be noted at this point that Keynes used the term ‘aggregate demand for labour’, and Darity and Young the term ‘aggregate demand schedule for labour’. The common usage of these terms (which is also suggested by Keynes’s distinction (1936, p.25) between aggregate demand and supply functions and the point of effective demand as the intersection of the two functions) would be the demand for labour in total. This is with labour treated as homogenous or measured in equivalent units as is confirmed by Keynes in the following quote:

‘For, in so far as different grades and kinds of labour and salaried assistance enjoy a more or less fixed relative remuneration, the quantity of employment can be sufficiently defined for our purposes by taking an hour’s employment of ordinary labour as our unit and weighting an hour’s employment of special labour in proportion to its remuneration; i.e. an hour of special labour remunerated at double ordinary rates will count as two units. We shall call the unit in which the quantity of employment is measured the labour-unit; and the money-wage of a labour-unit we shall call the wage-unit....This assumption of homogeneity in the supply of labour is

not upset by the obvious fact of great differences in the specialised skill of individual workers and in their suitability for different occupations.’ (Keynes, 1936, p.41)

Darity and Young (1997) argue that Keynes’s first definition of involuntary unemployment breaks down if the labour demand or labour supply schedules are sensitive to any endogenous variable other than real wages. We would endorse that view since we would argue that the demand for labour schedule is sensitive to the level of aggregate demand.

In order to consider the basis of the argument against Keynes’s first definition, we use the schedules for the ‘employment real wage relationship’ (ERWR) and the supply of labour, so as to illustrate how his definition relies on the slopes of these schedules being ‘conventional’. The ERWR, as shown above, is interpreted as portraying the link from effective demand to employment to the real wage. The supply of labour schedule is taken to reflect the maximum level of employment available at given real wages based on the idea that most workers face some compulsion to work, and that the available workforce is determined by certain social and institutional considerations regarding eligibility for work. Our underlying view of labour supply can be contrasted with the usual one (including that of Keynes) which rests on ‘disutility’ considerations, and which sees full employment as ‘choice-determined’ (for further discussion see Spencer, 2006; Sawyer and Spencer, 2006).

Figure 2 near here

In Figures 2a and 2b, the ERWR is portrayed as positively sloped. This is, of course, ruled out under perfect competition but under imperfect competition the real wage-employment relationship can readily be positively sloped over some range (as indicated above).

We start with Figure 2a where the ERWR is shallower than the upwards sloping labour supply schedule. If the adjustment process here were one of real wage changes, then this would be an unstable situation in that when the real wage is above the market clearing level, employment offered exceeds supply, and real wages would be presumed to rise. Consider

though the determination of the level of employment by effective demand. When employment is at L1, then there would be overfull employment, and a reduction in real wage would be associated with a decline in employment, and a decline in the extent of overfull employment. When employment is at L2, then there would be unemployment and a fall in unemployment would be associated with a rise in the real wage. Keynes's test for the existence of involuntary unemployment would fail in this case.

For Figure 2b, where the ERWR is steeper than the upwards sloping labour supply schedule, employment at L3 would involve unemployment (as S exceeds ERWR) and a reduction in employment would be accompanied by a reduction in unemployment (the gap between S and ERWR) and a reduction in the real wage. Employment at L4 would be one of overfull employment, and a rise in employment would be associated with a rise in the real wage. These results are again inconsistent with Keynes's first definition of involuntary unemployment.

In figures 2c and 2d both the ERWR and labour supply schedules are drawn as downward sloping. The shape of the labour supply schedule could be explained by the increased pressure to work that workers will face to meet their existing needs as wages fall: hence the view of a forward sloping labour supply curve at lower wage levels (see Prasch, 2000). Figure 2c implies that employment at L5 would involve overfull employment. Here, if we followed the letter of Keynes's first definition, we would have the false appearance of involuntary unemployment, in so far as an effective demand induced fall in real wages would call forth a greater increase in the available supply of labour (Robinson, 1937, p.172; see also Ambrosi, 2003, p.91). Employment at L6 would be compatible with unemployment, with an increase in employment being associated with a lower real wage and declining unemployment. Figure 2d would mean that at employment L7 there would be unemployment (since S exceeds ERWR) and a decline in employment would accompany a rise in real wages

and a reduction in unemployment, whereas employment at L8 would indicate overfull employment and an increase in employment would be associated with a fall in real wages and a decrease in the extent of overemployment.

In terms of Figure 2, there are four cases (L1, L4, L5, L8) where overfull employment arises under the assumption that employment is demand determined, and two of those (L1, L8) are cases where the real wage would be above the market clearing level. In two cases of unemployment (L2, L3) a decline in real wages would be associated with a decline in the demand for labour, and in two cases (L6, L7) a decline in real wages would be associated with demand for labour and supply of labour exceeding the initial level of employment. The clear point which emerges from these diagrams is that the relationships between real wages and employment, between real wages and unemployment and between unemployment and employment are model specific.

There may be a tendency to treat the diagrams drawn in figure 2 as a curiosum. But, in the context of labour, the widely-used idea of a backward bending supply of labour indicates the possibility that the supply of labour curve is negatively sloped, and imperfect competition (increasing returns) and some empirical evidence suggests that the relationship between real wage and employment is a positive one (e.g. Blanchflower and Oswald, 1994). These various diagrams serve to illustrate the extent to which Keynes's first definition of involuntary unemployment is model specific (in terms of the slopes of the schedules involved) and this definition would not identify involuntary unemployment in other situations.

Figure 3 near here

However, in Keynes's approach the marginal product of labour curve is not the demand for labour curve, and we now consider implications of that statement. In figure 3, suppose for a given level of effective demand, employment is at the level L^* . Whether or not L^* is a position of full employment depends on the real wage. In terms of Keynes's first definition of

involuntary unemployment, the real wage is at W_1 , and there is unemployment in terms of the difference between the available supply of labour (here denoting the maximum level of employment) and the effective demand determined level of actual employment. Like the neoclassicals, Keynes defined the labour supply curve in terms of the marginal disutility of labour (Spencer, 2006). If the real wage is at W_2 , then there would be a position of full employment in the sense of labour being on its supply curve (i.e. the real wage is equal to the marginal disutility of labour). There is then a sense in which unemployment in Keynes (1936) does arise from real wages being too high to clear the labour market but any real wage above the market clearing level denotes unemployment by definition – hence the association of unemployment with non-labour market clearing: however in Keynes (1936) there is no mechanism by which the real wage would fall from W_1 to W_2 . Although unemployment is perceived to be generated by a lack of demand (in the sense that demand is less than the equivalent of L_+), full employment could be restored by a drop in the real wage to W_2 (leaving aside issues such as the effect of real wages on demand and issues concerned with declining wages rather than lower wages).

Keynes's argument in chapter two of the *GT* proceeds as if the real wage is at W_1 , and his first definition of involuntary unemployment implies that the labour market features an excess supply of labour, with a rise in demand leading to a lower real wage, via higher prices. This will reduce involuntary unemployment through a combination of higher employment due to higher aggregate demand and increased voluntary leisure due to the induced fall in the real wage – the real wage adjusts in this case via changes in effective demand. The case of the real wage being too high is, of course, the flip side of prices being too low (relative to money wages). In effect, firms can be said to have too little market power to raise prices. In this situation, unemployment could be resolved if either firms received more market power and drove down real wages (clearly this would have negative effects on the distribution of

income), or if workers were not so keen to work as much, i.e. if the supply of labour curve shifted to the left.¹ In both cases, the change in unemployment can be seen to occur without any change in effective demand.

Keynes's first definition of involuntary unemployment suggests a change in the real wage as accompanying the fall in involuntary unemployment: but this fits in with later chapters of the GT only if the reduction in the real wage comes as a result of an increase in aggregate demand which generates an increase in output and thereby in marginal cost which raises the price of output, thus reducing the real wage. Further, it requires that firms have indeed set their prices (p) according to marginal cost (mc) (for then $p = mc = w/mpl$, w wage, mpl marginal product of labour, and hence $mpl = w/p$). If firms have set their prices above marginal cost, then the real wage is lower than otherwise and firms are not operating along the marginal product of labour schedule. In terms of Figure 3, firms could be operating at any real wage between $W1$ and $W2$. At real wage $W2$, there would be full employment in the sense of labour being on its supply curve. At a wage between $W2$ and $W3$, an increase in effective demand would increase employment without any necessary implications for the real wage: hence the great importance attached to aggregate demand in this case. A reduction in the real wage (when the real wage is between $W3$ and $W2$) would involve no change in employment though the quantity of labour supplied would be reduced. The existing level of employment would remain unchanged but full employment would diminish, due to an increase in voluntary leisure.

In this section we have given examples to support our contention that Keynes's first definition of involuntary unemployment is model specific and at most only applies where the 'demand' for and supply of labour have the conventional negative and positive slopes respectively. To what extent did Keynes move beyond the confusion created by the

discussion of unemployment contained in chapter two of the GT? This question is addressed in the following section.

4. Keynes's second definition

Keynes (1936) provided in chapter three of the GT a second definition of involuntary unemployment based on the responsiveness of employment to an increase in effective demand. This entailed defining involuntary unemployment in terms of its opposite, namely full employment. Thus, Keynes wrote, 'an alternative, though equivalent criterion [for full employment] is that at which we have now arrived, namely a situation in which aggregate employment is inelastic in response to an increase in effective demand for its output' (Keynes, 1936, p.26).

Darity and Young (1997, p.25) reproduce the following quotes from Keynes's writings after the GT which provide a similar view on full employment (and by implication on involuntary unemployment).

'The only reason why the orthodox theory denies the multiplier is because it is in fact assuming that there is always full employment, so that output as a whole has a zero elasticity' (Keynes 1973, p.58)

'[Full employment is defined] as the limiting case in which the supply of output ceases to be elastic ...' (Keynes 1973, p.85-6)

'If I were writing again, I should feel disposed to define full employment as being reached at the same moment at which the supply of output in general becomes inelastic' (Keynes 1973, p. 71)

'Indeed the condition in which the elasticity for output as a whole is zero is, I now think, the most convenient criterion for defining full employment' (Keynes 1973, p.101)

Keynes (1936, p.15) suggested that his two alternative definitions of involuntary unemployment amount ‘to the same thing’. Darity and Young (1997, pp.23-4) and de Vroey (2004, p.71) argue that the two definitions are, in fact, different. Keynes’s first definition can be interpreted as being consistent with the absence of labour market clearing, in which workers are ‘off’ their labour supply curve and firms are ‘on’ their labour demand curve. This is, in contrast, to the second definition that focuses simply on the elasticity of employment with respect to effective demand. Darity and Young comment ‘taken literally, this second definition is *not* identical with the first. The movement of real wages does not enter into this definition, nor, strictly speaking, does the market for labour. The definition says the economy is at full employment when no further increase in employment is forthcoming from an increase in aggregate demand’ (1997, p.24).

On the interpretation of Darity and Young, the most obvious reason why the elasticity for output as a whole will be zero is that ‘the economy has reached full capacity, given its prevailing stock of capital equipment and technical know-how’ (1997, p.25). Thus the direct link between the rise in aggregate demand and the increase in employment can be seen to depend on the extent of productive capacity carried by firms. Two comments can be made here. Firstly, it seems that even in chapter three of the GT Keynes believed that full employment would be defined by the labour supply curve based on the marginal disutility of labour. Hence his comment that ‘the volume of employment is not determined by the marginal disutility of labour measured in terms of real wages, except in so far as the supply of labour available at a given real wage sets a *maximum* level of employment’ (1936, p.30; emphasis in original). Full employment in this sense could be seen as set by subjective factors, independently of the level of productive capacity (see Spencer, 2006). The second comment relates to the possibility that full capacity working (of capital equipment) may be reached before employment reaches full employment, notably in the case where there is a

capital shortage. That is, employment may be unresponsive to increases in aggregate demand, even where the economy features positive levels of unemployment when there is inadequate productive capacity (which may include the distribution of productive capacity as well as the total level of productive capacity). We return to this issue below.

The second definition of involuntary unemployment given by Keynes (1936) refers to employment being *inelastic* to increases in aggregate demand, which could be read to mean that the proportionate rise in employment is less than the proportionate rise in demand. This applies in the (near to) full employment case, whereas employment is elastic with respect to aggregate demand in the case of involuntary unemployment. However, some of the quotes given subsequently by Keynes indicate a zero rise in employment following a rise in demand fits the case of full employment.

This second definition would seem to confront two difficulties. First, it suggests that either a position of full employment involves a vertical supply of labour curve or that a rise in demand has no implications for the real wage. When the supply of labour depends on the marginal disutility of labour, as Keynes himself assumed, then the supply curve may be positively or negatively sloped and a vertical curve would only arise as a special case (though Keynes himself only suggested a positively sloped curve). The supply of labour could then be expected to be sensitive to the real wage. Further, we are speaking here of the supply of labour in terms of hours and more specifically the hours individuals would like to supply: if we take the more realistic case where employers determine the hours of work, then the employees would not be able to actually vary their hours of work, and may be forced to work longer or shorter hours than they prefer. In chapter three of the GT, Keynes retained the neoclassical view that full employment (as defined by points on the preference-determined supply curve of labour) is compatible with a situation where all workers can decide their own hours of work (see Spencer, 2006). The second issue (that a rise in demand has no

implications for real wage) also runs counter to Keynes's argument that price varies with marginal cost and hence with output, and thereby the real wage varies with employment.

The second difficulty is that Keynes's second definition takes no account of the amount of productive capacity in existence. If firms are already operating with high levels of capacity utilisation, even in the face of high levels of unemployment, then increases in demand may not bring forward higher levels of output and employment. This can only occur, if there are net additions to capacity. The working assumption of Keynes appears to be that firms are operating with plenty of capacity, so that increases in demand can be absorbed by increases in output up to the point of full employment.² However, since Keynes maintained the assumption of perfect competition, the firms in operation are working in the range where the marginal cost is rising (to satisfy the second order profit maximising condition), and hence increases in output require some combination of existing firms expanding output with rising costs and other firms re-entering the market.

Figure 4 near here

A possible situation which fits this situation is given in Figure 4, with the ERWR drawn with a general form. The significant point here is that at a point such as A in Figure 4, any increase in demand would not lead to any increase in employment or output, but it would not be a position of full employment (as defined by labour being on its supply curve). Thus, it may be concluded that Keynes's second definition requires an assumption that, if aggregate demand is stimulated in the economy, there is sufficient productive capacity on which labour can be employed. In effect, it would be necessary to assume away any kind of capacity constraint on the achievement of full employment. Keynes's second definition provides a demand-deficient view of unemployment, but at the expense of overlooking unemployment which may arise from structural problems exhibited in a lack of productive capacity (relative to the potential workforce).

5. Imperfect competition

Darity and Young (1997, p.24) remark that ‘Keynes may have considered the page 15 and page 26 definitions as ‘equivalent’ because, at the time, it was inconceivable for him that employment and real wages could rise simultaneously’. Keynes (1939) seriously considered this possibility and began to revise his views on diminishing returns in response to evidence such as that of Dunlop (1939) and Tarshis (1939), but he never managed to revise his own theory of employment to take account of increasing returns. He did though note that on the basis of a positive relationship between employment and real wages that ‘if we can advance farther on the road to full employment than I had previously supposed without seriously affecting real hourly wages or the rate of profits per unit of output, the warnings of the anti-expansionists need cause us less anxiety’ (Keynes, 1939, p. 41).

Models of imperfect competition, from Chamberlin (1933) onwards (and in the macroeconomic context notably Weitzman, 1982) suggest firms operate with declining average costs, and rising productivity of labour. Figure 5 illustrates a possible situation where the ERWR is again the employment, real wage relationship and the rising part of that curve corresponds to declining costs. The level of employment in this approach would be set by the level of effective demand. Take an outcome such as point B. A move to full employment (at C) would require an increase in aggregate demand and an increase in the real product wage. In this situation, the two definitions of involuntary unemployment offered by Keynes would clearly conflict.

Figure 5 near here

The adjustment process in this situation is not one of the real wage changing (but if such an adjustment process were postulated then this situation would be seen as unstable, since real wages would tend to fall in the face of excess supply). The level of effective demand determines the level of employment, and that would change with decisions outside of the

'labour market'. It could be concluded from this that involuntary unemployment cannot be aligned with either of Keynes's two definitions: the page 15 one rests on particular assumptions on the slopes of the 'demand' and supply of labour functions, whereas the page 26 one rests on an assumption of excess capacity (and profitability). Relaxing either of these assumptions then requires that we search for an alternative definition of involuntary unemployment. We offer our own definition below.

6. The definition of involuntary unemployment

Involuntary unemployment can be readily defined as the difference between those who would be in paid employment in a situation of full employment and those who are actually in paid employment. A situation of full employment would be one where all who want or need to secure paid work are able to do so. This view of full employment recognizes that, whilst for some employment and the hours worked may be a matter of effective choice (in that the alternative is tolerable), for most it is not a matter of effective choice: to put it at its strongest if the alternative is no income and starvation, then there cannot be said to be a 'choice' not to work in any meaningful sense. The concept of unemployment is, in our view straightforward, and the main difficulty lies in translating that concept into an operational form.

There are two aspects of the concept of full employment to consider: what may be termed the social and the individual. Society through its laws and norms takes a view on who is regarded as in the workforce, and hence who would have to be employed if full employment were to be achieved; but also there are views on who is outside the workforce. The clearest example of this concerns children and the degree to which they are in the workforce and can legally take paid employment. There can be changes, sometimes abrupt, in who is regarded as in the available workforce: the war time experience in the UK being a prime example of an abrupt change in the attitude to the employment of married women. There have also been changing views on age of retirement, especially on early retirement (in the sense that early retirement is

a euphemism for unemployment, and requires to be supported by some part of society other than the individual in terms of payment of pension etc.).

It could be said that society deems certain groups to be outside the workforce: this may be undertaken through legal restrictions (as with child labour) or through the provision of some kind of income support (this may largely be through government, but also charities etc.). Within the group deemed 'eligible' for employment, some may have a degree of choice as to whether or not they offer themselves for employment. The very capital wealthy, for example, do have a choice as to whether to work or not. For most in this 'eligible' category, however, there may be perceived to be little or no choice, in the absence of other sources of income.³

Unemployment is by its nature involuntary. It is the situation in which some individuals, who prefer or are required to work, are not able to exercise that preference or satisfy that requirement. We have to recognise that any opportunities for paid employment may be heavily circumscribed and the individual has little option but to seek paid employment. Unemployment is then the situation in which the preferred state of employment cannot be attained. The state of involuntary unemployment has then to be distinguished from the state of leisure for the latter state is one which has been chosen by the individual concerned (in effect, because there is sufficient non-labour income for them to be able to make that choice). The notion and measurement of unemployment would be straightforward in a world where there was an unambiguous definition of the workforce (even if it is one which varied with the level of real wages etc.), where the labour force is homogenous and where there is no unemployment arising from mobility of labour between jobs. Let us look at these three conditions in turn.

- well-defined labour force

If the labour force could be defined with precision, then unemployment is simply the difference between the labour force and employment. By definition, unemployment is the

absence of employment where the latter is sought, and is thereby involuntary. Those who are not in paid employment and not seeking employment are not unemployed: they fit into a range of other categories; undertaking unpaid work in the home, retired, children, incapacitated from work through ill health etc.. Whilst the label of voluntary unemployed may be applied to some of these categories, it is a misuse of language. Thus, our notion of unemployment refers to those who would wish to have employment, those who would accept employment if offered and those who have to work but who are not currently looking for paid employment, say, for reasons of a lack of available jobs. This latter group could be seen as involuntarily detached from the available labour force, in so far as they would be actively searching for work, if they thought they had a chance of gaining work. Taking the category as a whole, many included in the above notion of unemployment would appear in the official statistical measures of unemployment (whether of the administrative form of those receiving unemployment benefits or of the survey form of responding to the question of whether you have sought work in the past few weeks and are available to begin work in the next few weeks). There would in general be some ‘disguised unemployment’ as well – that is people who in terms of our definition would be unemployed but who would not appear in the statistical measures of unemployment. When measured unemployment is based on a claimant count, the ‘disguised unemployed’ would include those available for work but not currently claiming or eligible for unemployment benefits. When measured unemployment is based on those who actively sought a job within a specified time period, the disguised unemployed would include those available for work but dissuaded from actively seeking work through perception of low probability of obtaining a job.

Some of the statistical measures of unemployment (e.g. those based on claimant count) may include a few who are not in paid employment but who are not available for work. This would not reinstate the notion of voluntary unemployment as a theoretical construct but

rather puts a label on those included in measured unemployed but who are not unemployed according to our definition.

- labour is not homogenous

If labour were homogenous (and abstracting from ‘frictional unemployment’) it would be relatively straightforward to say whether non-employment was or was not involuntary: if someone was willing to work at the going wage but was unable to find work, they would be unemployed. Someone who was not willing to work at the going wage would be outside the labour force. In this situation the only reason for retaining the term voluntarily unemployed may be to indicate that the person concerned would wish to supply labour at a higher wage (so in that sense is in the work force) but currently does not do so.

In a world where labour is homogenous in all aspects, then whether an individual was unemployed or not in a situation of overall unemployment would in effect be determined at random. Unemployment would be a negative ‘prize’ in a lottery. If labour were homogenous, then firms would have no way of choosing one person over another, and an individual worker could not do anything to enhance her chances of being employed.

However, in the real world, labour is far from homogeneous, and in a situation of demand for labour falling short of the supply, some rationing mechanism comes into play. Employers can use a range of criteria to decide who to employ and who not to employ. The characteristics of workers (which can range over many factors including skill levels, previous employment history, location etc.) may well influence who is employed and who is not. When some mechanism for the determination of who works and who does not exists, then the ‘sharpness’ of the notion of involuntary unemployment is lost at the level of the individual. Insofar as an individual can vary (or give the appearance of varying) characteristics (e.g. acquire skills, move location) which are relevant for the employment decision made by firms, then, at the level of the individual, there may be perceived to be some element of choice. An individual

can seek to acquire the characteristics which will enhance their employment prospects. If one's ability to push to the head of the queue is relevant, then those who do not push hard enough are the unemployed, and to some degree their unemployment has come from their own actions (or lack of actions). But, of course, a general increase in 'pushiness' does not reduce the level of unemployment. In a situation where there is a limited number of anything, then who gets what is available may to some degree be a matter of choice. Take the example of a highly popular sporting event where the attendance capacity is limited. Who is admitted involves some choice – those who are first to book, those who are prepared to bribe and cajole etc. get the tickets; for an individual there may be an element of choice; but collectively the number to be admitted is pre-determined. There are always some people who will be disappointed. Unemployment can be similarly viewed: the number of jobs is set by the level of aggregate demand, and how those jobs are allocated may depend on 'choice': it could be those with less work motivation are less likely to get a job, though having high work motivation would not be a guarantee for securing a job, given that the number of available jobs depends upon demand conditions in the economy. Indeed, where jobs are in short supply, due to low levels of demand, even the most highly motivated workers may face disappointment in obtaining paid employment.

From this perspective, the test of whether there is involuntary unemployment (as compared with individuals not in paid employment but who undertake other activities) would involve asking the following question: if one additional person became employed was this at the expense of someone else not being employed? Continuing the example given above, if one person gets a ticket for the sporting event has this replaced someone else getting one. It has to be recognized that this is a test which would be virtually impossible to apply in practice, but would seem to get to the heart of the notion of involuntary unemployment. It does not impose any particular notion as to why there is a lack of sufficient jobs to secure full employment.

This could arise from, *inter alia*, a lack of demand, a lack of productive capacity, or the presence of an inflation barrier. It also indicates that involuntary unemployment is a macroeconomic concept which does not have a precise individual counterpart. In effect, an individual may be regarded as voluntarily unemployed because they have not tried hard enough, searched enough, been prepared to take a low paid job, been unwilling to accept jobs involving much less skill than they have etc.. But if that individual tried harder, searched more, demanded lower wages, accepted low skilled work etc., in a situation of rationing there would be no effect on the overall level of employment since they would merely replace someone else. The level of involuntary unemployment would remain the same as well, only the membership of the unemployed pool would be altered.

- mobility between states

On the basis that the unemployed are those who wish to have work but cannot find it, the category voluntary unemployed in theoretical discussion disappears. It may make a reappearance in two contexts. First, when moving from theoretical discussion to measurement, some who are measured as unemployed may be deemed by some as not available for work and hence in that sense as voluntarily unemployed. But, in theoretical terms, there would be no difference between a rentier with sufficient income to pursue a life of leisure and a recipient of unemployment benefit who would not want to accept a job– both would be placed in our ‘other’ category and not be regarded as unemployed. People may well shift between categories over time, and as they do so the measure of unemployment would change: for example, the rentier who now decides to seek work but who is unable to find a job and the benefit recipient who responds to the perception of increased demand by seeking work would join those categorised as unemployed.

Second, the theoretical framework is essentially a static one and does not make any allowance for ‘frictional unemployment’ and the inevitable movement between jobs. In

general, 'frictional unemployment' may be regarded as involuntary unemployment at the level of the individual, and its extent is a function of the efficiency of the 'matching' process in the labour market. But, of course, the 'search unemployment' literature models 'frictional unemployment' in terms of decisions over the length of search in light of the incentives faced, and a longer period of search unemployment may be deemed as part of 'voluntary unemployment' by utility-maximising individuals.

At the level of the individual, unemployment may have an element of voluntariness in the sense that the individual could take actions which increases her chances of obtaining a job. But that would come at the expense of displacing some other worker. For workers as a whole, unemployment is essentially involuntary. In a world of homogenous labour, it is clear that unemployment is involuntary, as argued above. But there are systemic forces which determine the level of unemployment. Potentially these could be changed, and in that sense unemployment is a collectively chosen state.

What happens in the economy is the result of human decisions – there is nothing else. Hence if unemployment arises, there is some sense in which it arises from choice: collectively, humans could have made different decisions which would have led to full employment. But note if we view unemployment as endemic to capitalism, the choice involved would be an alternative economic system (as noted by Kalecki, 1943, 'fundamental reforms' would be required). If the central bank is viewed as suppressing demand to achieve an inflation target and unemployment results, when they could have taken the decision to raise demand levels, then unemployment results (in that context) as the decision of the central bank. Unemployment could then be said to be voluntary at the level of the central bank, but involuntary as far as the individuals who are unemployed.

Collective choices are important to the determination of unemployment. In particular, we would argue that the choice of policies to tackle unemployment is of vital importance.

Clearly, where the achievement of low unemployment is not afforded any priority, or where unemployment is viewed as necessary to achieve low inflation, the persistence of unemployment may be seen as the outcome of collective choices taken by policy makers (in this case, choices not to achieve full employment). We would argue that over recent years many governments and central banks in Western economies have taken the decision not to resolve the unemployment problem but instead have decided to prioritise the control of inflation, and that this has contributed to the persistent failure to achieve full employment in these countries.

7. Conclusions

This paper has focused on the definition of involuntary unemployment. The two definitions given by Keynes, it was argued, are best seen as tests of whether there is involuntary unemployment rather than as definitions of involuntary unemployment. Hence, if the experiment could be conducted of increasing aggregate demand and employment increased as a consequence, then that would be evidence of involuntary unemployment in the initial position. However, in this case involuntary unemployment (at the level of the individual) could exist and aggregate demand may not reduce it if there were inadequate productive capacity.

We have argued that unemployment as a theoretical construct is essentially involuntary. The term voluntary unemployment is a misuse of language. The appearance of voluntary unemployment arises only when the statistical measures of unemployment do not correspond to the concept of unemployment as discussed in this paper. Some individuals (e.g. the very rich) are in a position to choose between employment and 'leisure', but they are not choosing unemployment as we define it. We have also argued that involuntary unemployment has to be viewed as a macroeconomic concept rather than a microeconomic one. Within a given pool of unemployment, the characteristics, decisions and actions of individuals can influence

whether they are unemployed rather than others and in that sense have some appearance of voluntariness. But, in a collective sense, the degree of unemployment is not the outcome of decisions made by the individuals concerned. Rather, it is dependent upon certain other factors, most notably the level of aggregate demand. Finally, we would emphasise that some collective choice is always available to reduce unemployment and that governments can and should play an active part in securing full employment.

Endnotes

* We would like to thank Mark Hayes for his comments on an earlier version of this paper and we are grateful for the comments and guidance of the Editor of this journal. Remaining errors are our own.

1. Whether a particular level of employment is regarded as full employment here depends on whether that level of employment is on the supply of labour curve. In the neo-classical approach, a change in the 'taste' for employment would shift the supply curve and hence the level of full employment.

2. This was perhaps a reasonable assumption to make in the context of the 1930s depression, though even in these conditions, losses in capital and capacity through lack of replacement investment and idleness may still have meant that output and employment were slow to respond to rises in demand, in spite of high levels of unemployment.

3. We would note that self employment can act as one alternative source of income for those who are unemployed and we would include it as part of paid employment; however, we would stress that the viability of self-employment as a route out of unemployment will

depend on demand levels in the economy and the insufficiency of demand will prevent the unemployed becoming self-employed, even where the start-up costs of such employment are relatively low.

References

- Ambrosi, G.M., 2003. *Keynes, Pigou and Cambridge Keynesians. Authenticity and Analytical Perspectives in the Keynes-Classical Debate.* Basingstoke, Palgrave Macmillan.
- Blanchflower, D. Oswald, A.J., 1994. *The Wage Curve,* Cambridge, Mass.; London, MIT Press
- Chamberlin, E., 1933. *The Theory of Monopolistic Competition,* Cambridge, Mass, Harvard University Press.
- Clower, R.W., 1965. The Keynesian counter-revolution: a theoretical appraisal. in: Hahn, F., Brechling, F., (Eds.), *The Theory of Interest Rates.* London, Macmillan, pp. 103--25.
- Darity, W., Young, W., 1997. On Rewriting Chapter 2 of *The General Theory.* in: Harcourt, G.C., Riach, P., (Eds.), *A 'Second' Edition of The General Theory. Volume One.* London, Routledge, pp. 20--27.
- Davidson, P., 1983. The Marginal Product Curve is not the Demand Curve for Labor and Lucas's Labor Supply Function is not the Supply Curve for Labor in the Real World. *Journal of Post Keynesian Economics,* 6, pp. 105--17.
- De Vroey, M., 2004. *Involuntary Unemployment: The Elusive Quest for a Theory.* London, Routledge.
- Dunlop, J.T., 1938. The Movement of Real and Money Wage Rates. *Economic Journal,* 48, pp. 413--34.
- Kalecki, M., 1943. Political aspects of full employment. *Political Quarterly,* 14, pp. 347--56.

- Keynes, J.M., 1936. *The General Theory of Employment Interest and Money*. London, Macmillan.
- Keynes, J.M., 1939. Relative movements of real wages and output. *Economic Journal*, 49, pp. 34--51.
- Keynes, J.M., 1973. *The General Theory and After: Part II Defence and development: Collected Writings*, Vol. 14. London, Macmillan.
- Prasch, R.E., 2000. Reassessing the Labor Supply Curve. *Journal of Economic Issues*, 34, pp. 679--692.
- Robinson, J., 1937. *Essays in the Theory of Employment*. London, Macmillan
- Sawyer, M., 1992. Keynes's macroeconomic analysis and theories of imperfect competition, in: Gerrard, B., Hillard, J., (Eds.), *The Philosophy and Economics of Keynes*. Aldershot, Edward Elgar, pp.107--128
- Sawyer, M., 2002. The NAIRU, aggregate demand and investment. *Metroeconomica*, 53, pp. 66--94.
- Sawyer, M., Spencer, D., 2006. 'Labour supply, employment and unemployment in macroeconomics: a critical appraisal'. mimeo.
- Spencer, D., 2006. Work for all those who want it? Why the neoclassical labour supply curve is an inappropriate foundation for the theory of employment and unemployment. *Cambridge Journal of Economics*, forthcoming.
- Tarshis, L., 1939. Changes in Real and Money Wages. *Economic Journal*, 49, pp. 150--54.
- Weitzman, M., 1982. Increasing returns and the foundations of unemployment theory. *Economic Journal*, 92, pp. 787--804.

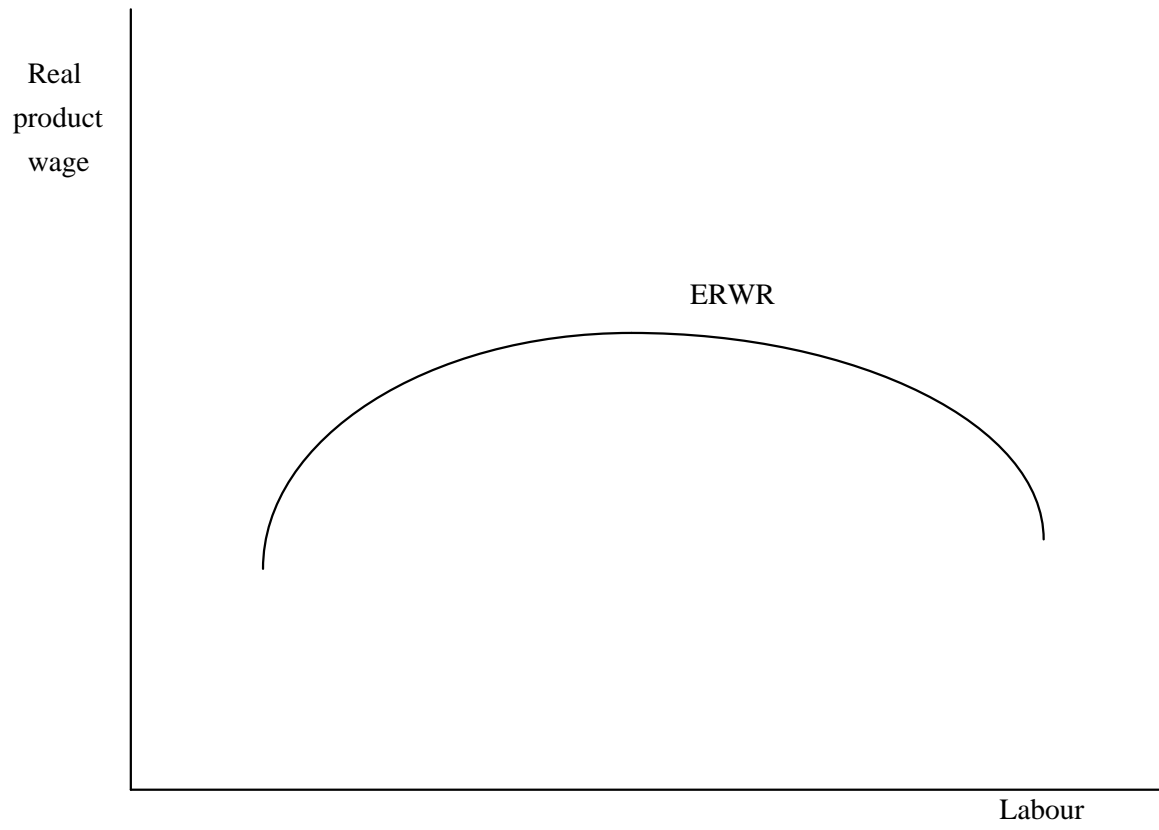


Figure 1 Real product wage-employment relationship

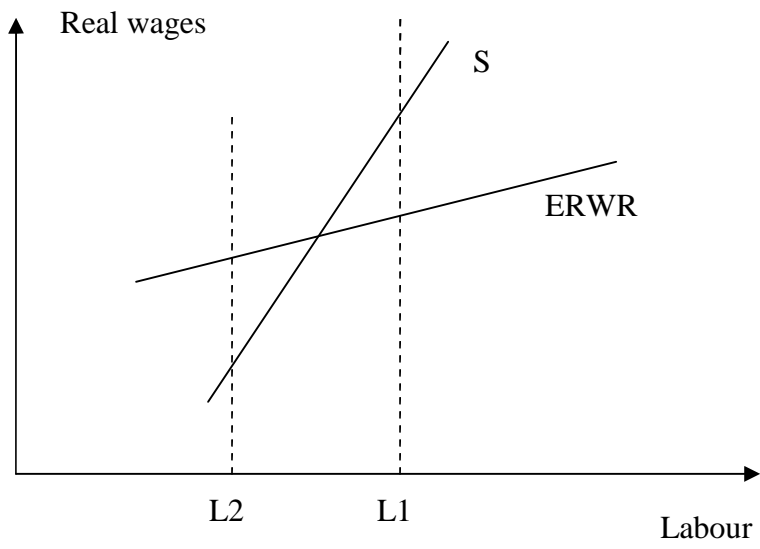


Figure 2a

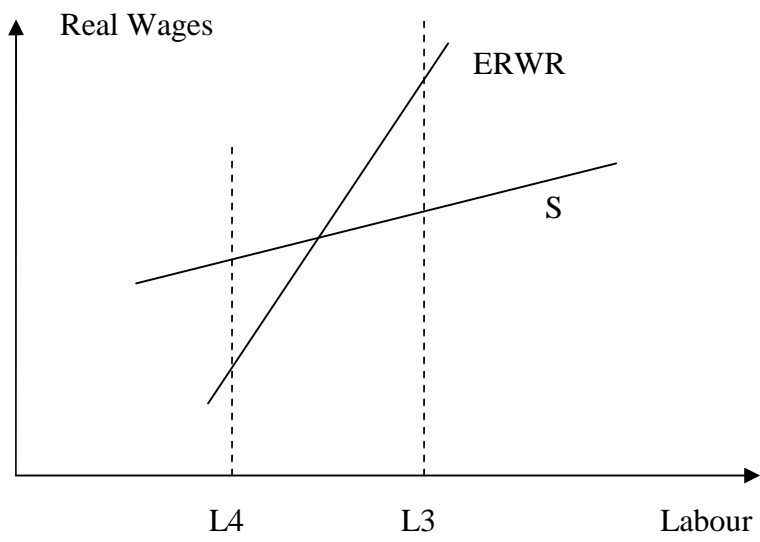


Figure 2b

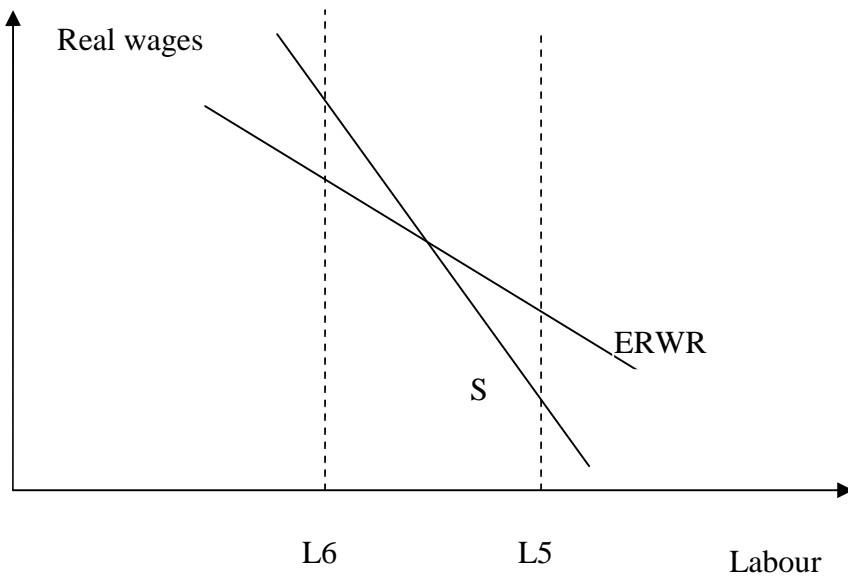


Figure 2c

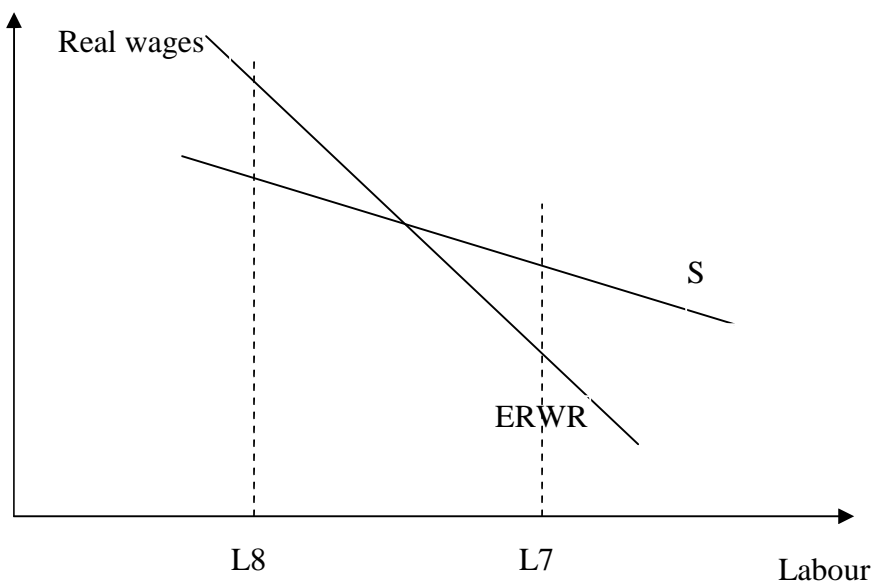


Figure 2d

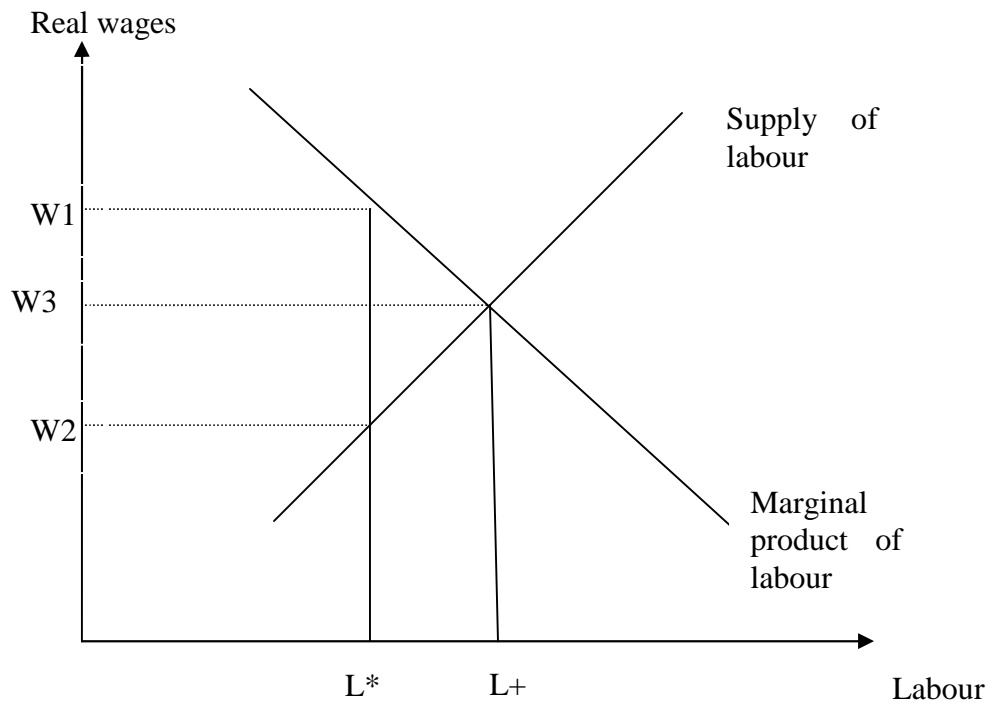


Figure 3

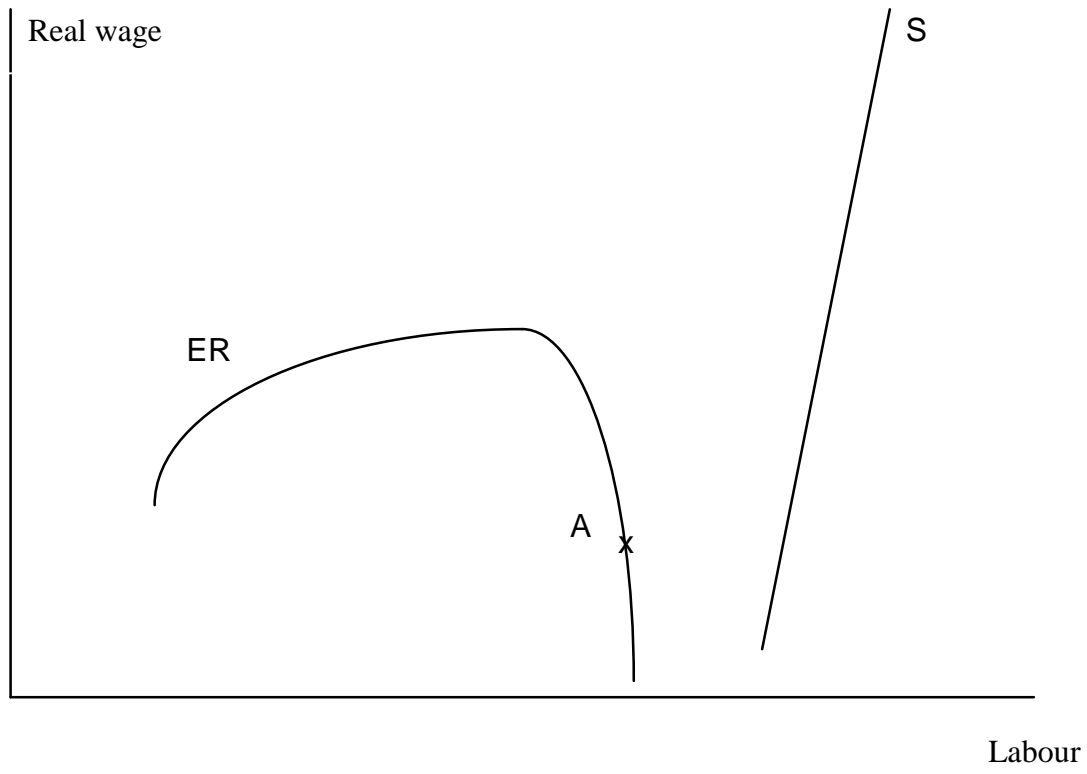


Figure 4

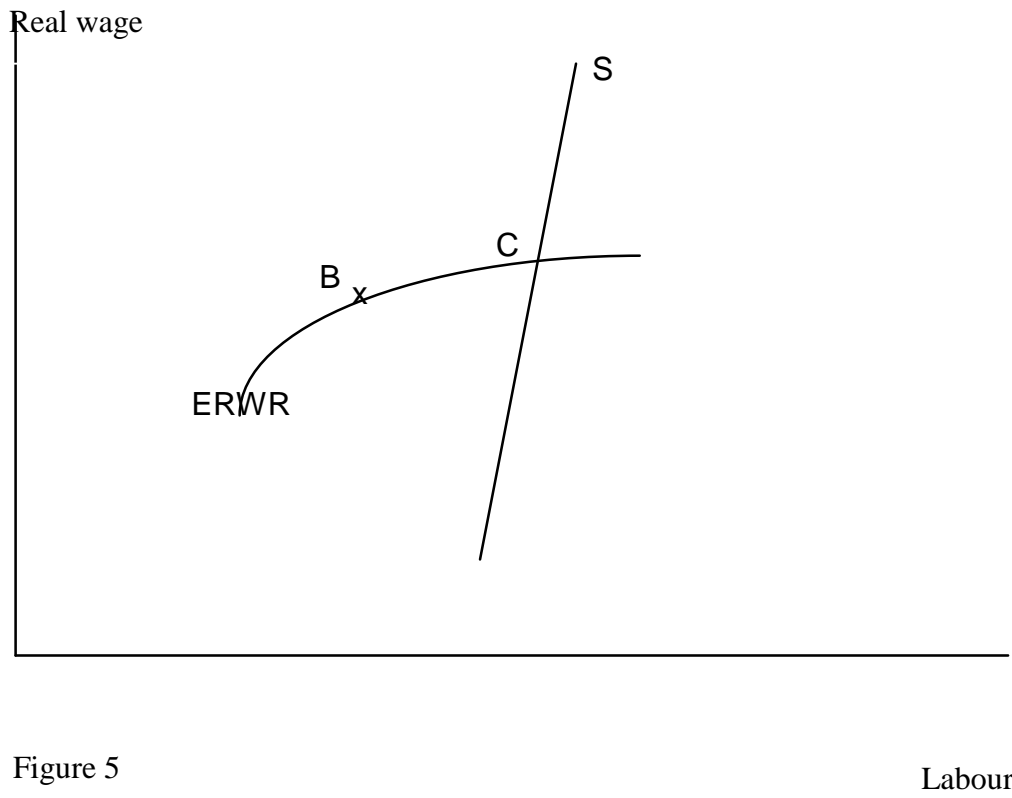


Figure 5