1978

Rigid Wages and the Adjustment Costs of the Firm

Albert Breton

Ronald Wintrobe

Follow this and additional works at: https://ir.lib.uwo.ca/economicsresrpt

Part of the Economics Commons

Citation of this paper:
RESEARCH REPORT 7829

RIGID WAGES AND THE ADJUSTMENT COSTS OF THE FIRM

by

Albert Breton and Ronald Wintrobe

October, 1978
RIGID WAGES AND THE ADJUSTMENT COSTS
OF THE FIRM

Albert Breton and Ronald Wintrobe*
1. INTRODUCTION

The issue of wage rigidity has been central to macroeconomic theory at least since the publication of Keynes's *General Theory* (1936) because, with fully flexible wages macro-theory predicts continuous full employment of labour. Wage rigidity is therefore the central assumption of theories of involuntary unemployment. No less central to macroeconomics, and partly the result of the pre-eminence imparted to it by Keynes's work, has been the tradition of seeking an explanation for wage rigidity on the supply side of the labour market. Whether he was assuming money illusion on the part of workers or advancing a variant of the relative-income hypothesis -- the wage differentials assumption -- Keynes always focused his attention on supply considerations.¹

Current work is still conducted in the tradition Keynes initiated. The most popular among current theories has been that of search, which explains wage rigidity by inserting the costs of acquiring information in the decision-making calculus of individuals. This theory has been faulted as an explanation of cyclical unemployment on the ground that it is implausible to suppose that information costs can account for the persistence of the unemployment that is sometimes observed. Other contributions have further weakened the theory by faulting it also as an explanation of frictional unemployment.

Among these is the work of Feldstein (1975) which has shown that a large fraction of job losses are the result, not of quits, but of layoffs and that most employees who do change jobs do so without any intervening period of unemployment, two results which contradict search
theory. A second contribution has been the incorporation, by Lucas (1972) and others, of the rational expectation hypothesis in macroeconomic models. This has led to the conclusion that search theory, as an explanation of wage rigidity, is weak on theoretical as well as on empirical grounds. At the same time, the recognition that wage and price flexibility coupled with rational expectations permit the derivation of most of the propositions of monetarism from otherwise simple Keynesian models has underscored once again the crucial importance of wage rigidity for an understanding of involuntary unemployment.

In response to these criticisms, recent efforts, pioneered by Azariadis (1975), Baily (1974) and Donald Gordon (1974), have sought to explain wage rigidity by appealing to the existence of contracts. For a number of reasons which we detail in the next section, we believe that this approach is also unsatisfactory. The difficulty economists have had with the concept of involuntary unemployment seems to be attributable to the fact that at the level of pure theory they have looked for an explanation of this phenomenon on the supply side of the labour market alone. Hence, it has been necessary to assume that workers, either individually or in the aggregate, somehow voluntarily agree to be made involuntarily unemployed.

Theories of rigid wages have focused on supply to the neglect of demand considerations for two further reasons which are grounded in the way the theory of the firm is usually formulated. The first is that at the level of the elementary textbook, it is assumed that competitive firms always take the price of output and of factors as given. In such
a world, it makes no sense to search for a hypothesis about sticky wages on the demand side of the labour market:

In more advanced theory, the search for an explanation of how wages adjust to inequalities of supply and demand has led to the view that competitive firms are transient monopsonists, while disequilibrium lasts and can, therefore, cut wages without losing all of their employees. However, it is usually assumed that firms which face a decline in the demand for their output and the consequent choice of maintaining stable wage rates and laying off workers, or of cutting wage rates and experiencing an equal or a smaller number of quits, must always prefer the latter course of action. The key assumption which leads to this result is that, even in advanced theories of the firm, there are no adjustment costs to wage rates reductions; more precisely, there is no rationale for adjustment costs to increase with the rate or speed of wage adjustment.

This paper departs from current theories in that it seeks to develop a theory of "sticky" wages and of involuntary unemployment based on the notion that it is rational for firms to respond to a decline in demand for their output initially by laying off workers rather than by cutting wage rates and experiencing voluntary quits, and by cutting wages only slowly over the longer run, if the decline in demand proves to be permanent. Wage rigidity, therefore, originates on the demand side of the labour market, in the adjustment costs to the firm of altering wage rates.

If one adopts this approach, several phenomena are immediately easier to understand. First, the costs of making wage changes provide a
rational explanation for the known fact that large firms and other hierarchal organizations change their wage scales only infrequently. Second, those costs explain why unemployed workers are not observed to seek employment by bidding down wages, since they know that their efforts would be spurned -- money wages being set by employers. Finally, and most importantly, the suggested hypothesis provides a logically satisfactory explanation for the existence of involuntary unemployment, as we shall seek to show.

Before proceeding with this task, we devote the next section to a demonstration that contract theory is not an acceptable alternative. Then, in Section 3, we outline the advantages of the approach taken here in contrast to the labour supply approach common to the search and contract theories and to earlier explanations of involuntary unemployment. In Section 4 we discuss in more detail the nature of the firm's adjustment problem. Section 5 examines the implications of the model for the theory of money wage rates and Section 6 concludes the paper.

2. CONTRACT THEORY

Contract theory is founded on two basic postulates. The first is that since the ownership of human capital cannot be directly traded in the marketplace, owners of human capital cannot diversify that component of their "portfolios" and therefore cannot reduce the risk of capital losses in the manner available to owners of non-human capital. Consequently, their marginal rate of substitution between risk and expected return will, ceteris paribus, be higher than that of stockholders --
namely, owners of non-human capital, in the present context, firms -- and therefore opportunities exist for "trade" between these two groups: owners of non-human capital can, as it were, sell insurance to owners of human capital. The second postulate is that this insurance is in fact provided to workers by stockholders via the institution of "implicit" or "quasi" contracts between workers and the firm. Wage rigidity follows from these two postulates because workers are held to prefer a fixed rather than a fluctuating price for their services as a consequence of their desire for security. Since workers are willing to pay for this security by accepting a lower average wage, it is in the interests of the firm to offer this package to them.

A first basic difficulty with this theory, as its originators themselves have clearly recognized, is the question of layoffs. Fixed wages imply fluctuating employment; therefore the question arises of why workers would enter into the type of contract that is postulated? In order to rationalize both wage rigidity and involuntary unemployment (i.e. layoffs), as a consequence of the assumed risk aversion of workers, contract theorists have therefore sought to show that these are somehow not the relevant choices facing individual workers. For example, Donald Gordon (1974) distinguishes two classes of worker -- tenured and non-tenured -- and assumes that tenured workers choose security, that is, choose to never be laid off, and therefore receive fixed wages and permanent employment. This leads to involuntary unemployment for the non-tenured workers, since they cannot bid down the wages of the tenured workers. Undoubtedly, this explains some of the unemployment in the real world.
But it does not explain the observed layoffs of tenured workers. In order to do that, Gordon has to introduce some element of risk-sharing between employers and employees. But a risk-sharing contract is consistent with almost any possible contractual arrangement between workers and the firm, and does not imply rigid wages.

Contract theory could therefore be rescued only if layoffs took place mainly among non-tenured workers, or were not very important as a source of unemployment. The evidence recently assembled by Feldstein (1975), however, has both dramatically confirmed the importance of layoffs in the United States at least, and shown that these are experienced by tenured as well as non-tenured employees. The evidence is that a very large fraction of all job losses are layoffs, not quits nor firings, that the typical period spent unemployed is very short (less than four weeks), and that most people (75%) who are laid off do in fact return to their former jobs. With respect to the cyclical component of unemployment, the data appear to show that temporary layoffs are an even larger fraction of the number of job losers. A larger fraction of unemployment is in fact experienced by workers who are "attached" to (have an implicit contract with) a particular firm but accept to be, and are, periodically laid off at their employers' discretion. It is this type of contractual arrangement which has to be explained by the new contract theory.4/

A second basic difficulty with contract theory is that, unless prices of commodities are assumed to be fixed, stability of money wages implies variable real wages in the face of price level changes. The
theory's first postulate -- risk aversion -- requires fixed real wages. The second postulate -- the existence of implicit contracts -- appears to require stability of money wages, at least for the duration of the contract. Moreover, it is stickiness of money, and not of real, wages which account for the observed existence of involuntary unemployment. But why would workers agree to rigid money wages in the interests of security, unless they were subject to money illusion, in which case, of course, we do not need the contract theory to explain unemployment. The logic of the contract theory must therefore be applied to prices as well as to wages, that is, both must be equally rigid and indeed, Okun's (1975) version of the theory runs along these lines. The theory's plausibility is, however, considerably weakened by this extension, in which "implicit contracts" explain rigidity of wages and prices in the same way that habits, frictions, lags and other such concepts do.

3. A NEW APPROACH

The model developed below is based on the notion that the division of production between markets and hierarchical organizations is not an accidental or haphazard development, but an optimum response to legal and property rights structures, to relative factor prices, to production technologies and to the technology of administration. The organization of production within the firm is similarly based on economic principles, and consequently these principles can be used to explain the structure of hierarchical organizations -- the size of firms, the extent of the division of labour within firms, the ratio of administrative
personnel to direct labour and so on.\textsuperscript{5/} The requirements of efficiency dictate that changes in technology, in demand conditions and in factor supplies affect output and factor prices in the long run very much as neo-classical theory predicts. The case is different, however, with respect to short run adjustments. The central argument of this paper is that the adoption by firms of relatively complex and structured hierarchical forms in the interests of long run efficiency affect the costs of alternative short run adjustment strategies, that is, dictate the particular paths of adjustment which firms can adopt in the short run. Our explanation of the short run stickiness of wages is based on two principles. The first is that firms are characterized by an internal division of labour, determined on the one hand by the gains in productivity which it makes possible, as discussed by Adam Smith, and on the other by the limitations imposed by the gains, which may be achieved through the market.\textsuperscript{6/} The principle therefore implies an intra-firm division of labour as well as a division of labour among firms.

The consequence of the division of labour within the firm is, in turn, two-fold. The first is that employees will be specialized to particular tasks and hence in the short run, not adaptable to different uses. Although non-adaptability of factors follows directly from the division of labour, the phenomenon will be magnified by bureaucratic costs of switching or redirecting labour to different uses, and it can be magnified further by the presence of union restrictions on the uses to which labour can be put. The second consequence of the division of
labour is that different specialized factors are complementary in the production of the final output of firms. Hence, they must be used in specific proportions to minimize costs. As is well known, deviations from these proportions will tend to increase average costs in proportion to the size of the deviation and to the degree of complementarity of the factors involved.

The second principle is that of uncertainty or imperfect information, which in the present context implies that firms do not know in advance the optimal system of wage cuts required to effect a desired reduction in factor employment. As a consequence, wage cuts in response to a fall in the demand for a firm's output always entail the possibility of "overshooting" the required cuts. We will show that both the frequency of overshooting, and, more importantly, the particular adjustment costs which accompany that overshooting are greatly magnified when a firm's internal division of labour is taken into account. The adjustment costs associated with layoffs are not affected in a comparable way.

On the supply side, we assume only that some workers prefer to search while employed; in contrast to the usual postulate of search models, workers do not all react to wage cuts by quitting immediately in order to search for a job elsewhere.

In the next section, we show that these three assumptions -- imperfect information, intra-firm division of labour, and on-the-job search by employees -- imply "sticky" wages.
4. THE ADJUSTMENT PROBLEM

To illustrate the firm's adjustment problem, consider a competitive industry which experiences a decline in the demand for its product. If the price of the output falls, each firm will want to reduce output, and will have to face the alternative of layoffs versus wage cuts to adjust to the new situation. To implement either strategy, the manager of a firm must first calculate the precise quantities of the different factors required to produce the new, lower, level of output at minimum cost. This information is relatively inexpensive, since once obtained, it may be used again and again subsequently.

If layoffs were chosen to implement the change in output, this calculation is all that is required. Consequently, if layoffs were always chosen, firms could draw up contingency plans for their personnel requirements at different levels of output, and thenceforth delegate the actual mechanics of such operations to lower level employees. Adjustment costs would be minimal, and the task of adaptation to short run changes in demand an essentially routine one.

Now consider the alternative policy of wage cuts. The basic problem with this policy is that, since the firm's employees are complementary, a set of wage cuts must be devised which induces factors to quit in the "correct" proportions. Optimal wage cuts must satisfy two conditions: (1) each marginal specialized factor must be paid the minimum wages required to keep it from leaving the firm over the relevant time period, and (2) those factors retained by the firm must remain in the proportions which minimize costs of the new (lower) level of output.
Since the supply elasticities of the different specialized factors will not, in general, be the same, or even related, no single wage cut will be adequate for the purpose of guaranteeing these two conditions. Moreover, the supply of all factors will be more elastic the longer the time they are given to adjust so that the ratio of long to short run elasticities of supply will vary between different factors. Consequently, no single system of wage cuts will achieve the intended result, and indeed, in any organization of reasonable size and complexity, the structure or pattern of wage cuts required will be exceedingly complex.

Do managers of firms possess the information required to implement such a system of wage cuts? Managers expect some employees to quit if their wages are cut, and they undoubtedly expect this number to be larger the larger the wage cut. But they do not know how many employees will quit for any given wage cut, as has already been pointed out by Robert Gordon (1976).

One reason that management cannot be presumed to know the optimal wage cuts in advance is that the optimal cuts required to implement any given reduction in personnel are continually changing. They change with every change in factor supply conditions and with every change in the wage policies of competitors. Consequently, information obtained from wage cuts implemented during a period of a decline in demand is, unlike the information on factor proportions, not useful for subsequent applications, that is, it depreciates quickly.

Managers therefore do not know the optimum levels of wage cuts and must discover them. Of course, the required information cannot be
obtained through casual inquiry ("how many employees will leave if we cut wages by 10%?"); the only strategy available is to cut the wage rate and observe the number of quits. If that number is relatively small, the wage rate can be reduced further, and if it is too large, the cut can be partially reversed until an equilibrium is eventually obtained.

The costs of experimenting with wage cuts are, however, peculiarly large as compared, say, with entrepreneurial experiments to find the maximum profit level of output. The reason is that there are special costs associated with excessive downward (as compared with upward) wage adjustments. To put it differently, the crucial difficulty with downward wage adjustment is the possibility of overshooting the optimal wage cut, and this possibility is obviously more likely, the faster the speed of wage cuts. To show the consequences of overshooting the optimal wage cut, assume that all cuts turn out to be correct -- except one, which turns out to be excessive -- so that the quantities of all factors required to produce the new desired level of output are correct except that there is a shortage of one complementary factor. The costs of this shortage to the firm are not merely equal to the foregone output of that factor, but to this plus the lost output due to the reduction in the productivity of all complementary factors. The total "adjustment costs" associated with any particular wage cut are then equal to the probability that that wage cut will turn out to be excessive, multiplied by the average total costs (loss in output) of overshooting. Since these costs are an increasing function of the extent of the shortage of any one factor (from the principle of diminishing marginal rate of substitution), marginal
adjustment costs are an increasing function of the extent of overshooting, and therefore of the speed of wage adjustment.

There is therefore, an asymmetry between the costs of excessive wage cuts, versus those of insufficient ones, since there are no repercussions to complementary factors from insufficient wage cuts -- only the costs of paying wages higher than necessary to keep factors from leaving the firm. This asymmetry, combined with the implication that the adjustment costs of excessive wage cuts increase more than proportionately with the magnitude of the excess, implies that the optimal wage cut for any factor falls short of the wage reduction which equates the supply and demand for labour, that is, wages are "sticky" downwards.

Put formally, the optimal wage cut for each factor is that which equalizes the marginal benefits from a reduction in the wages bill with marginal adjustment costs. The larger are these marginal adjustment costs, the greater the downward rigidity of wages.

The relative magnitudes of these costs and benefits are examined in the next section. Before turning to these and other implications of the model, we note that all three assumptions -- imperfect information, on-the-job search, and an intra-firm division of labour -- are necessary to obtain this result. The effect of imperfect information is obvious -- if firms knew the correct wage cuts in advance, no adjustment costs would arise with their implementation, since adjustment costs only occur in our model when the firm makes a mistake and cuts wages too much.

The effect of on-the-job search by workers is more subtle and provides an interesting contrast to the search model. In that model,
workers react to a wage cut by quitting their jobs to search elsewhere. They quit because, allegedly, they prefer to search for a new job while unemployed. If this postulated behaviour was in fact correct, the adjustment costs experienced by the firm in making wage cuts would be minor. For if workers reacted to wage cuts by quitting immediately, firms could easily learn the number of quits which would result from any given wage cut, and could therefore learn the size of the optimal wage cut. In other words, the firm could run the wage rate over any range in a relatively short period, observe the supply response, and thereby trace out the supply curve.\textsuperscript{9} If, on the other hand, workers do search while employed -- and the evidence indicates that this is to a large extent the case -- then the firm cannot quickly and easily learn what is the response to wage cuts; it must therefore either cut wages slowly over the longer run, or overshoot if it attempts to cut them quickly.

Finally, the adjustment costs of the firm would also be minor if there were no intra-firm division of labour, or if the factors employed by the firm were instantly adaptable to each other's tasks. For then the firm could maintain its production schedules unchanged during the period of factor shortage induced by the excess cuts. During the period in which the firm is losing employees as a result of the cut in their wages, it could redeploy its labour force appropriately with the same efficiency it normally possesses, so that days on which its labour force was too small could, subject to capacity limitations, be compensated for at other times.
When factors are complements and non-adaptable to different uses, this cannot be done. On any day in which factors become available in the "wrong" proportions, the shortage of any one type of job-specific factor will reduce the output of the entire organization. This means that wage cuts which overshoot the optimal cuts for any one highly complementary factor will result in large permanent losses of production, and these losses will persist until the wages for that factor have been increased sufficiently to enable the firm to rehire the appropriate number of personnel.

5. THE THEORY OF MONEY WAGE RATE ADJUSTMENTS

The basic implication of the model is that the benefits of wage reductions (i.e. the size of the reduction in the wage bill) must be compared with their costs in order to decide on the appropriate speed of adjustment. Thus, for example, the longer the expected duration of any decline in demand, the larger are the benefits of wage cuts, and consequently the faster the optimal speed of wage adjustment. On the other hand, firms in industries where shifts in demand are relatively frequent will tend to use layoffs rather than wage cuts to adjust in the short run, and to cut wages relatively slowly over the longer run if the decline in demand proves to be relatively permanent.

The model also implies that ceteris paribus the costs of adjustment for any firm will tend to vary directly with its size, with the extent of its division of labour, with the elasticity of substitution between the different types of labour employed by the firm, and inversely
with the costs of retraining employees to adapt to each other's tasks. Since the speed of wage adjustment varies directly with the costs of adjustment, these factors determine both the relative preference for layoffs versus wage cuts in different industries, and differences in rates of change of wage rates for given changes in demand in those industries. Some evidence that these rates do vary considerably has been assembled by Hall (1975); evidence that they do vary is, of course, consistent with our model but inconsistent with the search and contract theories.

Thirdly, the model implies that the speed of money wage adjustment will be larger the larger the excess supply of labour. This holds for two reasons. Firstly, the probability that any given wage cut will overshoot the required reduction is smaller, the larger the wage cut required to reach the new equilibrium. Secondly, the costs of overshooting will be smaller when there is considerable excess labour supply, since workers will be both less likely to quit and more easily replaceable if they do quit. Thus the model implies a classic short run Phillips relation between the level of excess supply of labour and the rate of change of wages, where it is indeed money wages, and not real wages that should be on the vertical axis. Of course, as long as the supply of labour depends on real wages, this Phillips curve will shift up and down with changes in inflationary expectations, just as the conventional one does.

Fourthly, there is an asymmetry in the model between the costs of upward and those of downward wage adjustments. No difficult personnel adjustment problems are encountered by the firm in raising wages to meet
a situation of excess demand, since the firm can simply cease hiring when all its vacancies are filled. Nevertheless, wages will not be perfectly flexible upward, because even if the firm is unlikely to overshoote its personnel requirements, it can overshoot the wage increase necessary to meet them. The firm will therefore not wish to raise wages too high in the short run for the simple reason that it may thereby put itself in the position of having to cut them later for long run equilibrium purposes.

It follows that firms faced with an increase in demand for their output will not, in the short run, wish to expand production sufficiently to meet the entire increment, since to do so would mean raising wages beyond the level which will be required in the longer run. Hence the firm may rationally permit short run shortages to appear and to persist, rather than incur the adjustment costs of cutting wages at a later date. This implies that the short run supply curve will be steeper than the long run curve, even when no diminishing returns are present.\(^{10/}\)

Nevertheless, an asymmetry in adjustment costs between wage increments and wage cuts is obviously present: indeed, wages will tend to rise faster in response to an excess demand for labour then they will fall in response to an equal size excess supply. Moreover, for any given level of unemployment, the accompanying rate of wage inflation will be higher if that level is falling than if it is rising; in other words, the Phillips curve will exhibit anti-clockwise "loops". Again, this conclusion can be derived from the fact that the movement of the unemployment rate in either direction reduces the probability that wage rate changes (of the opposite sign) will overshoot the required changes.
Finally, it should be emphasized that the present model, unlike previous explanations of wage rigidity, is capable of explaining why the extent of wage rigidity in an economy tends to change over time, and why wages tend to be more rigid in some economies than in others. That is, the present model provides a theoretical foundation for the explanation of long run shifts in the Phillips curve which are not simply due to changes in inflationary expectations, but come about as a result of changes in an economy's industrial structure. For example, long run increases in the average size and structural complexity of firms imply an increasingly rigid wage structure, and hence it is easy to understand why wages tend to be more rigid in contemporary industrial economies than they were a century ago.

6. CONCLUSION

A theory of short run stickiness of money wages which is consistent with empirical findings and which also possesses solid microeconomic foundations is necessary to explain involuntary unemployment. Previous efforts can be faulted on one or the other of these two counts. In the foregoing pages, we have suggested a theory of wage stickiness which originates on the demand side of the market, in the costs to firms of downward wage adjustments. We have argued that these costs increase sufficiently rapidly with the speed of wage adjustment to explain the short run rigidity of wages.
NOTES

* Professor of Economics, University of Toronto and Assistant Professor of Economics, University of Western Ontario. We would like to thank Alan Hynes, Yehuda Kotowitz and Frank Reid for their comments. All remaining errors, of course, are ours.

1/ For a recent discussion of what Keynes was arguing, see Trevithick (1976).

2/ Arrow (1959) is still one of the best formulations of the problem.


4/ The hypothesis suggested by Feldstein in a follow-up paper (1976) to explain his numbers is, as the reader can easily verify, not inconsistent with the view of the short run adjustment mechanics for the firm which we are advancing here.

5/ See Wintrobe (1977) for a detailed model along these lines. In addition, much, somewhat unsystematic knowledge on these issues has been assembled by sociologists and organization theorists. See March (1965). For other economic approaches, see Alchian and Demsetz (1972) and Williamson (1975).

Complementarity may be defined in several ways, not all of which are mutually consistent. See the discussion in Samuelson (1974). The best definition for our purposes is that which distinguishes complementarity from substitutability in production by the sign of the cross-partial derivative of output with respect to different inputs, i.e. inputs are complements or substitutes as \( \frac{\partial^2 Q}{\partial x \partial y} < 0 \), where \( Q \) is output and \( x \) and \( y \) are inputs.

It may be asked why we do not simply equate the total loss in output from a marginal reduction (shortage) in one factor with the marginal product of that factor. Although marginal products can be defined this way, and to do so raises no difficulties with the present model, the more usual procedure is to identify the marginal product of a given factor assuming not only fixed quantities of other factors, but also that the other factors are perfectly malleable. Thus, in Robertson's example (1946), if we start with 10 men and 10 shovels, and then withdraw one man, his marginal product is computed as if the 10 shovels are magically transformed into 9 better quality shovels. This is the procedure followed here. The additional loss in output due to the fact that, in the short run, other factors are not perfectly adaptable or malleable, (in the short run the 10 shovels cannot be transformed into 9 better shovels and one of them is simply left idle) is referred to in the text as a loss in the productivity of complementary factors. For further discussion of these problems, see Robertson (1946) and Stigler (1939).
9/ A similar argument is made by Gordon and Hynes (1970) p.374 with respect to prices.

10/ This argument is analogous to that of Lucas (1967).
REFERENCES


_____________, "The Division of Labour is Limited by the Extent of the Market", Journal of Political Economy, 59 (June 1951) pp. 185-193.

