A brief survey on the role of trade unions in labour markets

Adriana Cassoni
**Brief historical overview**

The analysis of the economic effects of trade unions on labour markets has evolved in a quite unsteady way, with empirical findings posing theoretical unsolved puzzles and theoretical analyses being either almost absent or amazingly prolific (Pencavel, 1991).

In the industrial relations literature many hypotheses on the incidence and structure of collective bargaining were developed a long time ago. These, along with some standard economic theory analyses, motivated quite an important amount of empirical research. At this early stage and until the late sixties, economic theorists generally considered trade unions as institutions that would shift the labour supply curve but would not influence the competitive characteristics of the labour market.

The fact that the wage level observed was different depending on the presence or absence of a union stimulated an upsurge of empirical research on the effects of unions on wage setting. Thus, this early work was devoted mainly to explaining or verifying the existence of a wage gap between unionised and non-unionised firms, industries and/or countries (see Lewis, 1986 for a survey). In doing so, researchers were faced with the dilemma of defining an adequate utility function for the union to maximise, of choosing the appropriate decision variables and, to a smaller extent, of finding a suitable way of including the characteristics of workers and firms among the determinants of unionism. However, as Johnson (1975) noted, there was still no consensus on unions' goals while theoretical foundations were still missing. Moreover, despite the many case studies which were carried out, various methodological issues were not addressed during this period, such as how critically the models relied on the exogeneity of the variables involved or whether it was possible to obtain similar results with different assumptions on the utility functions and bargaining processes. Despite these shortcomings, the analyses brought out many stylised facts that fed into the theoretical research.

One of the first aims of economic theorists was to explain the bargaining processes by which outcomes arose and to discuss the optimality or efficiency of these outcomes. However, for this to be done rigorously, the analysis of unions' preferences had to be thoroughly revised so that the variables included in the bargaining agenda could be clearly determined. Although researchers have been discussing the issue for decades (Dunlop, 1944 and Ross, 1948 are early references), there is as yet no consensus on unions caring or not about the level of employment. Empirical evidence is inconclusive.
as, although observed contracts do not generally include clauses on employment, some studies have shown that employment does matter to unions (Farber, 1978; Dertouzos and Pencavel, 1981; Carruth and Oswald, 1985 are examples).

In any case, the first models developed assumed that bargaining took place only over the wage level, thus implying that management determined employment unilaterally according to the labour demand curve. Two models became quite popular, depending on the role assigned to management in the negotiations: if it was to take the wage as given -set by the union- then the monopoly union model (the origin of which is attributed to Dunlop, 1944) was proposed, while if it was assumed to bargain with the union to determine the wage level the right-to-manage model was considered as an appropriate description (Nickell, 1982). The models yielded similar predictions, although in the latter the resulting wage level was lower than in the former, as firms would have some bargaining power. However, both implied that the outcome was Pareto non efficient, that is, that the wage-employment pair observed would generally be worse, for at least one party, than other possible outcomes. This gave rise, once again, to the debate on the inclusion/exclusion of the employment level as a union goal, although this time for theoretical instead of empirical reasons.

The above argument led to the formulation of the efficient contracts model (MacDonald and Solow, 1981), in which unions and management were assumed to bargain simultaneously over both wages and employment, so that Pareto optimality was guaranteed. The wage-employment outcomes predicted would lay on the contract curve, to the right of the labour demand. However, the model would still be inconsistent with the observed fact that negotiations do not generally include employment explicitly in the bargaining agenda (see Oswald, 1993 for a recent extensive survey).

Many empirical studies were carried out during this period trying to assess the reliability of the different formulations in explaining real phenomena. Moreover, some researchers tried to select among them testing their validity for specific data sets (Brown and Ashenfelter, 1986; Card, 1986; among others). However, the procedures used have been recently criticised because of not being robust to changes in the underlying assumptions (Pencavel, 1991; Manning, 1994). Hence, the adequacy of the models should be seen as a major empirical research topic, as the implications of the alternative formulations are very different. Firstly, the efficiency or inefficiency of the outcome is important in terms of the behaviour of agents. To assume that the parties ignore a potential gain as is the case in
the right-to-manage and monopoly union models, could cast doubts on the adequacy of the bargaining process proposed. Secondly, small wage increases and/or low wage levels could be attributed to the union caring about employment or to the firm rejecting the union's proposed wage and succeeding. Again, the above implies different objective functions and bargaining frameworks. Finally, the efficient contracts model predicts a higher level of employment over the cycle than that resulting from the right-to-manage and the monopoly union formulations, in which employment is, further, unaffected by union "bargaining power" over this variable.

During the second half of the seventies and the early eighties, the debate on union preferences and goals as well as on which of the two formulations was more suitable -both theoretically and empirically- was still widespread. One approach to reconcile the assumption of efficiency with the unobserved bargaining over employment was developed in the implicit contracts theory. According to this view, uncertainty on the economic environment to be faced and the assumption that workers need to have an insurance against 'bad' states of nature, would promote the design of contracts contingent on the future economic performance, unknown ex ante but observed by both parties ex post (Baily, 1974; Azariadis, 1975). Thus, the risk aversion of the parties would justify that they accepted a contract in which they would share the risk. Further developments proposed employment contingent contracts (Calvo and Phelps, 1977). Assuming that the firms would possess better information than workers about the state of nature and that workers would never observe the value of certain variables, the only way to prevent the firm from cheating once the wage was set, would be to link the wage to the employment level. Alternative formulations based on private information further added some insight to the approach (see Oswald, 1986b). Despite the resulting contracts were optimum, the approach was criticised mainly because real contracts are generally very simple while contingent contracts would be not only complex to design but very difficult to monitor (Oswald, 1986a). Further, as wages are generally settled for a fixed period of time while employment is not, there would always exist an incentive for the firm to default. Some authors argued, however, that the existence of a union itself as well as the repeated nature of union-management relationships could be enforcing the contract (Malcomson, 1983).

A second line of research restated the union's preferences. Two main arguments were put forward. First, if the implicit rule for firing workers was "first in, last out", then seniority would be an important variable to be taken into account in specifying union's preferences. Under the new scheme it was shown
that it was possible to obtain the efficient outcome at the point in which the marginal revenue of labour equals the wage (Grossman, 1983; Oswald, 1985; Oswald, 1993). An analogous result was obtained if the utility function of the union was specified differently depending on employment being smaller or greater than membership. Once it was considered that the union's concern over employment would disappear when all members got a job, the union's indifference curves would become horizontal and, again, it would be possible that the tangency between isoprofit contours and indifference curves would be lying on the labour demand curve (Oswald, 1985; Carruth and Oswald, 1987).

Another proposed way to overcome the obstacle of imposing a priori restrictions on the structure of bargaining was to include other variables in the agenda, that would be negotiated separately of the wage. Thus, although actual contracts do not include any statement on employment, it could be thought that indirect agreements on this matter would be negotiated in further stages -such as the assignment of workers to machines- or that wage bargaining might be done at a centralised level while employment and/or other issues are negotiated at the industry or firm level. At each stage the parties might have different bargaining power, due to the union having an unequal interest on the variables or an unbalanced capacity of summoning members for different issues. Models with these characteristics could be labelled multistage or sequential bargaining models (Manning, 1987; Card, 1990; Johnson, 1990). Their main consequence is that if the wage is set in a first stage and employment in a second one, the model would nest the three previous formulations and empirical tests on the significance of the parameters could be used as a means of identifying the bargaining procedure that is relevant for a specific data set. The outcome predicted by the multistage model would be somewhere between the labour demand curve result and the contract curve outcome depending on the relative bargaining powers. Thus, from a theoretical point of view, efficiency would be possible but is neither imposed nor subject to a specific utility function. Still, the possibility of agents cooperating to obtain a non optimum employment-wage outcome under certain circumstances would remain. However, since the difference in union power at various stages is the origin of inefficiency, the model might suggest a distinct direction of analysis related to the determinants of union's strength.

Finally, an alternative approach was to view the bargaining process as a repeated instead of a one shot game (Espinoza and Rhee, 1989). The repeated interaction among the parties would allow for reputation effects as well as for the possibility of their being punished in the future when deviating from previous agreements. As a consequence, incentive compatibility would be guaranteed and it would be possible
to observe an efficient outcome without bargaining over employment. The right-to-manage and the efficient contracts models would be special cases of this general formulation, depending on the time preferences of the parties and the discount rates they use.

In the second half of the eighties, attention was also drawn to other aspects of the economics of trade unions. The aim was basically to analyse how the predictions of the by then standard models would be affected by the relaxation of some of their simplifying assumptions. Among them, the hypothesis of fixed membership was revised in different ways. One example is Sampson (1988) in which membership is endogenous and where uncertainty is also incorporated by allowing for a stochastic labour demand schedule. Other researchers addressed topics such as the effects of the size of union membership; the free-rider problem -enjoy the benefits without incurring the costs- as well as the implications of heterogeneous members (Booth, 1985; Naylor, 1989; Booth and Ulph, 1990). Further, for the empirical research to better approximate real world processes, it was necessary to modify some aspects such as, for example, the static nature of the models. The inclusion of dynamics was justified in terms of the existence of employment adjustment costs relative to negotiations themselves, coordination, adjustment to shocks, etc. (Lockwood and Manning, 1989 is an example) and also on the endogenisation of membership (Kidd and Oswald, 1987; Lockwood and Manning, 1987).

Lastly, more attention started to be paid to the role played by unions in wage formation at a macroeconomic level. Thus, the analyses tried to shed some light on the way that unions' reaction to fiscal policies would influence their formulation as well as on how the policies themselves would restrict unions’ claims. The bargaining over wages was hence analysed as a game between a centralised union and the government, the result of which would be considered as an explanation of the unemployment rate evolution in previous decades. At the same time, decentralised bargaining was also studied, trying to evaluate the consequences of this different structure on the macroeconomic regularities observed (Calmfors, 1985). Calmfors and Driffl (1988) addressed the issue of how the above two different settings would influence macroeconomic performance.
The state of the art

Theoretical research on the economics of trade unions is quite profuse nowadays despite the apparent decline of these institutions all around the world. One of the main reasons for this is the renewed interest in how their presence and strength, as well as their structure and coverage, can affect the expected results of economic policies. In a world where deregulation, liberalisation and integration are being largely discussed and implemented with very different results, the role of the diverse institutional settings arises as a main issue.

Are the structure and coverage of unionism as well the degree of co-ordination in bargaining, features that help explaining the unequal performance of different economies? Is it that economists have not yet found an adequate model to represent the trade union's goals and the bargaining process? What are the consequences for investment, innovation and total factor productivity of the diverse institutional settings? How can the presence or absence of trade unions as well as their structure be related to the evolution of employment and wages over the business cycle? Are they helpful or harmful when firms and economies are subject to exogenous shocks? These are some of the questions that economists are attempting to answer by improving both the theoretical framework and the quality of empirical studies.

The concerns of theorists in the nineties might be divided in two broad categories. Firstly, there is a considerable amount of papers that continue and deepen the analysis of some of the topics brought up by the previous literature. Secondly, a renewed interest in the critical analysis of the structure of bargaining models is being addressed.

Regarding the first class of work, four main topics can be identified. Firstly, centralised versus decentralised bargaining. Although some work on corporatism had been done by the end of the eighties, Calmfors and Driffl's (1988) paper gave rise to a huge amount of empirical and theoretical research on how the structure of negotiations affects the outcomes. This proved to be a neglected area of analysis and to have important consequences for theoretical and empirical research which, up to that moment, had relied on the assumption of centralised or decentralised bargaining (see Calmfors, 1993 for a survey). The topic is being currently related to coordination and synchronisation issues as well as to multi-unionism (Naylor, 1995) and multi-level bargaining (Calmfors, 1993 and references therein). Further, a
lot of attention is being paid to how different structures might influence the expected results of macroeconomic policies such as rising the degree of openness, deregulating the labour market or implementing tax reforms (Driffill and van der Ploeg, 1993; Heylen, 1993; Rama, 1993a, 1993b, 1994).

A second field is that of private information and the role of strikes. This stream of research is the natural extension of some of the work carried out in the eighties in which strikes were considered a form of signal that would enhance credibility in bargaining when there are informational imperfections and sequential bargaining (Chatterjee and Samuelson, 1987; Hart, 1989 examples). Some of the recent analyses develop various sophistications, such as time-varying threats (Cramton and Tracy, 1994); the consequences of centralised and decentralised bargaining with respect to strike frequency and duration (Goerke, 1994); the relationship between costs of strikes and their length (Card and Olson, 1995); the effects of signalling and the incentive structure on the outcome of standard models (Vetter, 1995).

Thirdly, considerable attention is being devoted to models in which membership is an endogenous variable. This issue had been mentioned since the very beginning because many results depended critically on this assumption. An early theoretical example is the equivalence of the utilitarian and the expected utility functions only if membership is fixed (Oswald, 1985). Regarding empirical work, simultaneity and exogeneity biases may arise if the hypothesis of given membership does not hold. As was mentioned in the previous section, in the eighties there were some attempts to include membership dynamics and to analyse its implications for the models. Currently, the main ideas are related either to the existence of an intertemporal objective function that links actual employment to future membership (Jones and McKenna, 1994) or to the simultaneous determination of membership and wages (Booth and Chatterji, 1993 and 1995).

Lastly, the effect of unions on some specific variables such as productivity, innovation and hours of work has gained considerable attention in the past few years. Regarding productivity, it is argued that opposed to the monopoly "face" of unions, they possess a productivity enhancing facet. This is related to higher morale; the availability of public services, such as better job conditions or improvement in information channels; the involvement of workers in the performance of the firm; the existence of a link between workers and management without the fear of retaliating against those who complain; among others (Rosen, 1989; Wadhawani, 1990; Moreton, 1993).
Regarding innovation the main question posed is if because of the fear of lower employment, unions would be opposed to R&D activities and the introduction of new technology (Ulph and Ulph, 1988 and 1994; van Reenen, 1995 are examples). Again, the analyses show that the structure of bargaining will have a decisive effect on the final result. Finally, regarding hours of work, the existence of fixed costs and different adjustment costs in the context of a dynamic analysis implies that hours and employment should be treated distinctly by the firm. Further, unions have probably different preferences among various combinations of number of jobs and hours worked by employees. Hence, the inclusion of hours in the objective function of the union might have interesting consequences for standard models (Earle and Pencavel, 1990; Oswald and Walker, 1994).

The second main branch of current research analyses the weakness, or lack of robustness, of standard models when faced with changes in the underlying assumptions, such as those related to the variables used as an indirect means of bargaining over employment or to the production function (Clark, 1990; Johnson, 1990; Layard and Nickell, 1990; Manning, 1994; Benassy, 1995). However, there have also been remarkable advances by modelling the union-management bargaining process as a repeated game, so that the current behaviour of the parties would build a reputation and, if it is the case, would imply agents will be subject to punishment in the future (Haltiwanger and Harrington, 1991; Kandori, 1991; Rotemberg and Woodford, 1992). The existence of a punishment for those who deviate from equilibrium strategies could make contracts incentive compatible, and thus equilibrium may depend, for example, on the discount rates of the agents (Espinoza and Rhee, 1989). This, in turn, may be use to understand the behaviour of employment and wages in the business cycle (Schultz, 1994). Uncertainty enters naturally in this setup, as expectations on the future economic performance might have a crucial role in determining their willingness to cheat, the credibility of their threats and the plausibility of their offers.
References


__________ and D.T. Ulph (1990), 'Union wages and employment with endogenous membership', University College London, Department of Economics, mimeo.


J.T. Dunlop (1944), Wage determination under trade unions, MacMillan Ed.


L. Goerke (1994), 'Asymmetric information, centralized bargaining and strike outcomes', Universitat Hamburg, mimeo.


J. Haltiwanger and J. Harrington (1991), 'The impact of cyclical demand movements on collusive


_________ and R. Jackman (1990), Unemployment: macroeconomic performance and the labour market, Oxford University Press.

H.G. Lewis (1986), Union relative wage effects: a survey, Chicago, IL, University of Chicago Press.

B. Lockwood and A. Manning (1987), 'Dynamic wage-employment bargains with endogenous membership', Birkbeck College, mimeo.


___________ (1986a), 'A theory of non-contingent wage contracts', LSE, Centre for Labour Economics Discussion Paper 266.


C. Schultz (1994), 'Wages and employment in a repeated game with revenue fluctuations', University of Copenhagen, Institute of Economics, mimeo.


J. van Reenen (1995), 'The creation and capture of rents: wages and innovation in a panel of UK companies', University College London Discussion Paper 1017.
