The economics of the trade union

ALISON L. BOOTH
University of Essex

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1995
Reprinted 1996

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

Booth, Alison L.
The economics of the trade union / Alison L. Booth.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0 521 46467 6 (hc) – ISBN 0 521 46839 6 (pbk)
1. Trade unions – Economic aspects – Great Britain. 2. Trade unions – Economic aspects – United States. 3. Labor market – Great Britain. 4. Labor market – United States. I. Title.
HD6664.B598 1995
331.88´0941–dc20 93-50224 CIP

ISBN 0 521 46467 6 hardback
ISBN 0 521 46839 6 paperback

Transferred to digital printing 2002

SE

Contents

Figures

Tables

1 Introduction

Over the past two decades, there has been an extraordinary expansion of the economic theory of the trade union. There is also a huge empirical literature describing and quantifying the impact of the trade union on a host of labour market outcomes.[1] The purpose of this book is to impose some structure on this body of literature, in order to make it easily accessible to the student of the economics of the trade union or the economics of industrial relations. The book is not intended to be a comprehensive survey of the literature. Rather, it aims to provide a selective review of the crucial features of the analytical framework used by economists in modelling trade unions, a framework that is also potentially applicable to nonunionised labour markets where workers have some market power. Certain issues have been deliberately omitted in the interests of keeping the book to a manageable size for the readership for which it is intended. The book aims to be accessible in the main to third-year undergraduates and MSc students in economics, as well as to general readers with some basic training in economics and quantitative techniques. The main technical parts of the analytical framework have been confined to appendixes where appropriate, in order to avoid distracting the reader's attention from the main themes.

The book is written on the assumption that the reader is familiar with the perfectly competitive model of the labour market, in which labour is treated essentially as a commodity in a spot market. It is also assumed that the reader is aware of the weaknesses of this model, and in particular of the features of labour that distinguish it from other inputs, and which mean that labour cannot be treated in quite the same way as other factor inputs (Marshall, 1948). The two principal distinguishing characteristics of labour are, first, that workers retain ownership of their human capital (in the absence of slavery), and, secondly, that workers must be present at the workplace for the delivery of their skills. The fact that workers retain

[1] See, for example, references in the surveys by Oswald (1985), Pencavel (1985), Farber (1986), Hirsch and Addison (1986) and Ulph and Ulph (1990).

ownership of their human capital has the implication that any education or skills associated with employment are the property of the worker, who can therefore exercise some control over the use of these skills, and perhaps extract any surplus associated with them. The fact that workers must be present for the delivery of their skills means that workers must live near the workplace, which may constrain the opportunities of other family members, and make workers vulnerable to monopsonistic behaviour. It also means that the social aspects of the work environment are important for the worker.

There are a number of other important characteristics of labour; the differences between capital and labour with respect to these characteristics are sometimes of degree only, for capital may share some of these characteristics. An obvious feature of labour is that it is perishable; if a worker's skills become obsolete, the time and cost of training may mean that retraining is uneconomic. Moreover, the time and cost of training for workers, whose lifetime is finite, preclude labour from diversifying its portfolio of human capital. This specialisation of labour in one particular skill type, which may become obsolete, may leave the worker vulnerable to exploitation by the employer. The worker may, however, have some bargaining power through the fact that it takes time to provide additional supplies of specialised labour. Moreover, new employees may be imperfect substitutes for incumbent workers, for example where the incumbent worker has better information than management about work effort and production at the shop-floor level. In such a situation, the worker may refuse to reveal this information to management or to a new employee unless it is in his or her interest to do so, a fact that gives the incumbent some bargaining power. In general, in circumstances where it is costly for the firm to replace existing workers by outsiders for whatever reason, the incumbent workers have some market power.

Another important characteristic of labour is that the production process involves the bringing together of workers, giving greater opportunity for collective action. This also facilitates the development of personal relationships between workers, and sometimes between workers and management, which obviously does not happen with physical capital. Some of these characteristics of labour are conducive to collective action by workers, for the production process typically involves grouping workers together in a situation where there may be opportunities for exploitation by agents – either management or workers or both – of any surplus that might arise at the workplace. Where both workers and management have some bargaining power, bargaining may be either between individuals and management, or between an agent of the workers and management, where the agent could be a trade union. While bargaining may be more effective between a trade union and management, there may still be scope for individual bargaining with management in some production processes, owing to the presence of labour turnover costs, which give workers some monopoly power.

The models in this book refer to labour contracts between a trade union and management. However, these models are also relevant to a much wider class of situations than those in which a trade union explicitly represents workers, and have been extensively used in the insider–outsider literature (see Lindbeck and Snower, 1988) and the implicit contract literature. Individual incumbent workers in labour markets characterised by labour turnover or transactions costs possess a degree of monopoly power which, through individual or group bargaining even in the absence of trade unions, might be used to induce rent-sharing by management. Indeed, union contracts may be viewed simply as an explicit formulation of a wider variety of labour contracts that are found in labour markets. It is the explicit nature of union collective agreements that has allowed empirical research on union wage and employment determination, an avenue of research that is rather more difficult where bargaining agreements are not clearly specified in generally accessible written agreements.

The focus of the empirical work referred to in this book is primarily on Britain, but reference is also made to some major empirical findings from US studies. The institutional structure, and the stylised facts, associated with trade unionisation to which the book refers, relate both to Britain and the USA. However, there are considerable institutional differences between these two countries, which are highlighted in chapter 2 and also, where appropriate, in discussion of the theoretical models. Many of the innovations in the microeconometric modelling of union effects first emerged in the USA, where large cross-section and panel survey data at the individual level have been available since the late 1960s. For this reason, when discussing stylised facts or empirical regularities associated with trade unions, we sometimes refer to pioneering US work.

Much of the *theoretical* modelling referred to is appropriate to the unionised economies of Europe, Australia and New Zealand, and to the unionised sectors of the USA. None the less, it must be emphasised that there are important institutional differences across countries. In the introduction to chapter 2 we pay particular attention to the institutional differences in collective bargaining between Britain and the USA. It is important that students and researchers are aware of these differences, in order to avoid the pitfall of inappropriate applications of theoretical models characterising the behaviour of unions in one particular country to another country with a different institutional structure.

A reader might observe that there is some irony in the fact that,

Table 1.1 *An international comparison of union membership as a percentage of all employees 1970–85*

Country	1970	1975	1980	1985
Australia	51	54	52	51
Canada	36	34	30	31
France	23	23	19	na
Germany (FR)	37	39	39	37
Holland	38	39	37	29
Ireland	52	53	55	46
Italy	33	42	43	40
Japan	35	34	31	29
New Zealand	40	43	46	41
Norway	62	61	65	na
Sweden	74	79	88	88
United Kingdom	49	51	53	44
United States	27	25	23	16

Notes:
(i) 'na' denotes no data available.
(ii) Membership figures for the US are from the Bureau of Labor Statistics (see the notes to table 2.1 for discussion of US data sources for membership).
(iii) Break in the continuity of measurement in Canadian data in 1978 causes a fall of approximately 4 percentage points.
(iv) See Price (1989) for extensive discussion of data sources and data problems.
Sources: Price (1989) and Brown and Wadhwani (1990).

concurrent with the recent blossoming of research on trade unions, there has been a decline in trade union density in most major industrialised economies. Table 1.1 shows union density in eleven countries for the period 1970–85, where the definition of density is trade union membership as a percentage of the employed workforce. The table shows that density has declined over the period 1980 to 1985 for most countries, the exceptions being Canada and Sweden where density has remained constant. The USA had the lowest union density in 1985, with just 16% of all employees being trade union members.

Given that union density appears to be declining in most industrialised countries, and that in some countries only a minority of the workforce is unionised, do we really need to worry about providing appropriate models of trade union behaviour? For example, in the case of the US where union density is only 16%, might we be better advised to adhere to other theories of wage setting and worker behaviour? At the start of this chapter, the generality of union bargaining models was emphasised. While the models in this book refer to explicit labour contracts between a trade union and management, these models are also relevant to a broader class of situations than those in which a trade union explicitly represents workers.

Moreover, trade union influence in a particular economy extends beyond the direct measure of union power suggested by the union density figures. Industrial relations and labour researchers in the USA, where union density has been declining since the mid-1950s, have for a long time suggested that union influence extends well beyond the unionised sector. For example, the threat of union organisation of a nonunion sector may provoke management to provide wages and working conditions that mimic those negotiated in union firms. The idea is that the nonunion workers, therefore, will be less prone to unionise, since there is little difference between their welfare in a union firm and in a nonunion firm providing matching benefits. In addition, the threat of union organisation may provoke management into directing resources into anti-union activities, and resource allocation in the non-union sector will therefore be indirectly affected by trade unionism. Furthermore, modelling the behaviour of the trade union in partially unionised economies is obviously of importance for sectoral analysis of the parts of the economy that *are* heavily unionised, or where a powerful sector is unionised and there are knock-on effects for the rest of the economy through particular institutional aspects of wage setting.

For European countries and Australia and New Zealand, the influence of trade unions at the macroeconomic level is perhaps better indicated by the extent of coverage of the workforce by union collective agreements, rather than by the measure of union density. According to Layard, Nickell and Jackman (1991: 52), all European countries with the exception of Switzerland have over three-quarters of employees covered by collective agreements, as does Australia. Using data for full-time employees in a permanent job from the 1991 British Household Panel Study (BHPS), I have calculated that 57% of British male workers are covered by collective agreements, while only 44% are trade union members. The comparable figures for women are 57% and 40% respectively. Part of the difference between Layard *et al.*'s (1991) figures for UK coverage and the 1991 BHPS figures arises from the fact that Layard *et al.* include Wages Council workers in their 'covered' category. Of the major industrial countries, only in the USA are less than one-quarter of workers covered by union collective agreements. In 1991, union membership density in the USA stood at 16%, while contract coverage stood at 18% of employed civilian wage and salary workers (Hirsch and Macpherson, 1993).[2] This suggests that an understanding of the economics of the trade union is vital for analysis of the

[2] These figures are calculated using data from the 1991 Current Population Survey.

workings of the macroeconomy for most industrialised countries, and that at the microeconomic level it is also important for the USA.

In order to understand unionised labour markets, it is desirable to construct analytical models of trade union behaviour and test these against the data, and also to measure the impact of unions on economic performance. The theoretical analysis of trade unions has, to a large extent, proceeded separately from empirical measurement, and there is a far larger literature devoted to the latter. Indeed, there is an enormous body of empirical research on trade unions; in particular, on the observed differences between union and nonunion wages, and, to a lesser extent, on union and nonunion sectors, jobs and outcomes. This is particularly the case for the USA, where many of the results have been surveyed in Freeman and Medoff (1984), Lewis (1986), Hirsch and Addison (1986) and Addison and Hirsch (1989).[3] Much of the literature measuring the impact of trade unions on various outcomes has been largely descriptive with no theoretical foundation. But our knowledge as to why there are differences between union and nonunion labour markets is incomplete. As Farber noted in 1986, a comment that is still relevant today,

> there is quite a bit of controversy about what these differences mean. Are they accurate measures of the *effects* of unions, are they biased estimates of the effects, or are they statistical artifacts? How can these estimates be used to predict union response to changing economic conditions? Without a complete understanding of union behavior and how the outcomes of collective bargaining are determined it is difficult to answer these questions. (Farber, 1986: 1040.)

In essence, while descriptive empirical research is instructive in summarising data sets and suggesting stylised facts that require explanation, care should be taken in using it to make inferences about the impact of trade unions, as will be emphasised in the chapters in this book covering descriptive empirical research on union effects.

In chapter 2, we shall be examining the institutional structure of unionism and collective bargaining in both Britain and the USA. Examination of the trade union and its collective bargaining framework is the primary focus of the discipline of industrial relations, but it is also of crucial concern to economists wishing to model the behaviour of unionised labour markets. Trade unions in Britain and the USA are institutions that have evolved over several hundred years. To understand their structure, it is necessary for the student to have some knowledge of their historical development and the forces that led to their assuming a particular institutional form at the present time. Moreover, an understanding of their historical development can shed some light on the economic factors

[3] For a survey of Australian studies, see Miller and Mulvey (1993).

associated with union emergence, existence and stability, which will be helpful to the discussion of trade union theories in later chapters. Such an understanding can also illuminate current-day phenomena, such as British attitudes to European Community legislation on labour market issues. Owing to space limitations, the historical outline of chapter 2 is confined to a broadbrush approach, including only those factors that are believed to be relevant to explaining the present collective bargaining framework, as well as union emergence, existence and power.

The standard view of trade unions is that they are organisations whose purpose is to improve the material welfare of members, principally by raising wages above the competitive wage level. There is little dispute that unions are frequently able to push wages above the competitive level – what is called the 'monopoly' role of trade unions. There is an enormous body of literature documenting the impact of unionism on wage gains. Some of this will be examined in chapter 6. There is also a somewhat smaller body of literature examining the impact of unions on other variables, such as wage dispersion, productivity, profitability, investment and employment. Some of these studies will be examined in chapter 7. All of this literature is focusing on finding an answer to the important question: what do unions do? and thereby providing stylised facts or empirical regularities associated with trade unions. To the extent that these studies provide stylised facts associated with unionism, the reader familiar with the accepted economic methodology might expect an outline of these empirical regularities before the derivation of the theory.[4] However, since some of the more recent of these studies are based on the theory of trade unions developed in chapters 3–5, it seems more appropriate to place the empirical work aiming to measure trade union effects *after* the development of the theory. We therefore defer a review of empirical measurement of union effects until after a broad theoretical framework has been presented in those chapters.

Chapter 3 sketches what have become known as 'the two faces' of unionism – the monopoly role of unions, and the possible efficiency-enhancing role (Freeman and Medoff, 1979). The monopoly face of unions

[4] The accepted methodology of economics is that, following observation of an empirical regularity, economists can formulate a theory to attempt to explain the stylised fact. For simplicity and tractability, the theory necessarily abstracts from reality. In so doing, it makes simplifying assumptions. From the theory can be logically derived predictions for the economy or for particular economic variables. These predictions can then be tested against the available data. Thus the theory is inspired by empirical observation, and its predictions are also tested against the data. Since such testing is carried out with data sets with limited numbers of observations (rather than against the whole population), the theory can only ever be accepted or rejected probabilistically. It would be fair to say that few economic theories of the trade union have yet been adequately tested against the data. There are some notable exceptions to this, which will be discussed in association with the relevant theories in later chapters.

represents the orthodox textbook approach to unionisation, and is summarised in the first part of chapter 3. However, this chapter also provides a broad theoretical overview of the economic theory which attempts to answer some of these questions about how unions obtain the power to extract wage increases. Some of the theory is, as yet, undeveloped, and the chapter necessarily provides only a sketch of the major issues. Other parts of the theory have been well developed, in particular the analysis of union objectives and union–firm wage determination. Some of this latter work, now widely used in the literature, is expanded more formally in chapters 4 and 5. Because chapter 3 provides only an overview of the theoretical issues, the use of algebra is kept to a minimum, with broad discussion of the important issues and diagrammatic exposition where appropriate.

Chapters 4 and 5 show the formal derivation of the theoretical models that have achieved orthodoxy in the economics literature, and with which all labour economics students need to be familiar in order to read journal articles in the field. Chapter 4 considers ways of modelling the differing objectives of both the trade union and the firm, and examines the simplest method of reconciling the objectives of trade union and management: the monopoly union model. This method does not rely on any bargaining between the two parties. Instead, it supposes that the union is able to act as a monopolist in the supply of labour, and to impose a wage rate on the firm. The firm, however, retains its managerial prerogative to determine the number of workers to employ at the union wage rate. All workers are typically assumed to be identical in this type of model. We do, however, also examine the median voter model of trade union behaviour, where workers are assumed to be heterogeneous; the membership implications of a simple version of this model are considered.

Chapter 5 examines more sophisticated methods of reconciling the differing preferences of union and management, using a bargaining approach. Bargaining models of strike behaviour, and union–firm bargaining over standard hours, are also examined. The class of models developed in chapters 4 and 5 is widely used in theoretical work, and also in some empirical work aiming to estimate wage and employment equations based on a theoretical foundation. Emphasis is placed in these chapters on the simplifying assumptions underlying these models, the relaxation of which may well provide scope for future research.

With a broad theoretical framework in place, chapters 6 and 7 examine the impact of unions in Britain and the USA, and discuss which union effects appear to have the status of stylised facts. Chapter 6 focuses on the union–nonunion wage differential, which has often been used as a measure of union power. Macroeconomic models and models of the aggregate labour market commonly use the union wage markup for this purpose, since it is argued to be positively correlated with union power (Layard and Nickell, 1985, 1986). In chapter 3 it is argued that, at the microeconomic level, the existence of economic rents is a necessary condition for union wage markups, but that higher union wages will be found only where the trade union has the power to force the firm to give up some of its surplus. Therefore, the union wage markup will be positively correlated with union power, as suggested in macroeconomic work. But calculation of the union wage differential is sensitive to the method of estimation and the degree of aggregation of the data. Estimates from aggregate data are typically much larger than those from individual level survey data, which are in turn larger than estimates from panel studies. Which estimates can we believe? Is it a stylised fact that unions raise wages by 70% as one British researcher has claimed, or by just 7% as claimed by another? What are the problems involved in the calculation of these estimates? To understand the issues involved, it is important that we understand the methods used in estimation of union wage markups, and the econometric problems facing researchers in this field. To this end, chapter 6 examines just a few empirical studies which have attempted to quantify the impact of unions on wages, with discussion of the econometric issues throughout. These are presented non-technically in the main, so that the argument can be understood by a reader who has followed a standard econometric course.

A condition for a union to achieve wage gains is that the union has the necessary power to force the firm to share any surplus with the union; an alternative is that the firm may be willing to grant higher wages in return for increases in productivity that increase the surplus available to the firm. While the 'monopoly face' view of trade unions stresses the negative aspects of unionism, an alternative view is that in some circumstances they might be productivity-enhancing. Although unions may cause wages to increase in the union sector, neither employment nor firms' profitability need necessarily be greatly affected, since firms' higher labour costs may be offset by improved productivity. Since there are a variety of theories suggesting opposing union effects on productivity, it is ultimately an empirical issue as to whether unions are associated with increased or decreased productivity. Chapter 7 examines some empirical studies estimating union effects on productivity and productivity growth, investment, profitability, hours and employment.

Chapter 8 of the book turns to macroeconomic issues, in particular, the modelling of trade unions in the aggregate labour market. The chapter examines how the formal union models of chapters 4 and 5, which are microeconomic in flavour, might be applied to the macroeconomy. Over the past decade, macroeconomic modelling has been shifting increasingly from the Walrasian market-clearing approach to one in which account is taken of the fact that agents in the economy may be able to act strategically. The 'New Keynesian' approach to macroeconomics emerged in the 1980s as a response to the inability of the received macroeconomics to explain the phenomena of the period. By the 1980s, there were changes in the performance of the major advanced economies compared with the 1960s and 1970s. Growth almost halved, unemployment increased dramatically through both the 1970s and 1980s, and inflation accelerated in the 1970s, falling back in the 1980s to a level that was, none the less, higher than in the 1950s and 1960s. Moreover, while the OECD countries generally had a poorer economic performance in the 1970s and 1980s, there were considerable differences among countries with respect to growth, unemployment and inflation. These changes contributed to a re-evaluation of macroeconomic modelling, of which the 'New Keynesianism' emerging in the 1980s was one outcome; this approach is based on imperfect competition and combines Keynesian features with equilibrium unemployment. It attempts to obtain price and wage stickiness through modelling the imperfectly competitive behaviour of firms and workers, thereby providing microeconomic foundations for the behaviour of macroeconomic aggregates such as output, unemployment and inflation.

Chapter 8 examines the contribution of the recent microeconomic trade union literature to the New Keynesian approach to macroeconomics. The framework used in this chapter draws in particular on the approach of Carlin and Soskice (1990), Layard *et al.* (1991) and Blanchard and Fischer (1989). The chapter examines the implications of trade unions for wage and unemployment determination at the aggregate level, and draws principally on the monopoly union model of chapter 4. Of course, equilibrium unemployment may also be affected by the collective bargaining structure. Chapter 8 therefore examines the implications of the collective bargaining structure – in particular, the degree of centralisation of wage bargaining – for aggregate models of union behaviour and wages, prices, and unemployment. The implications of the macro-model for unemployment persistence, or hysteresis, are also considered, as is the validity of the model for economies which are only partially unionised.

The final chapter of the book draws some conclusions, and highlights areas where further research would appear to the author to be desirable. Students who are interested in further reading in any particular areas are referred to the references given throughout the book, and in the case of the extensive US empirical literature to the survey works of Freeman and Medoff (1984), Hirsch and Addison (1986) and Addison and Hirsch (1989).

While union membership and power have declined in many developed countries in the 1980s, it is unlikely that trade unions will vanish from European labour markets. Moreover, while the models in this book refer to explicit labour contracts between a trade union and management, these models are also relevant to a much wider class of situations than those in which a trade union explicitly represents workers. Individual incumbent workers in labour markets characterised by labour turnover or transaction costs possess a degree of monopoly power which, through individual or group bargaining even in the absence of trade unions, might be used to induce rent-sharing by management. The type of analysis used in modelling explicit labour contracts in unionised labour markets is also potentially applicable, with a few minor modifications, to these nonunion labour markets where incumbent workers have some market power.

The economic analysis of unionised labour markets has progressed enormously over the past two decades. However, there are omissions and oversimplifications in the literature, some of which are pointed out in the book. Therefore, much work remains to be done by the next generation of researchers in this interesting field.