Research Reports
Kansantaloustieteen laitoksen tutkimuksia, No. 77:1998
Dissertationes Oeconomicae

MARKUS SOVALA

Studies on wage formation
Studies on wage formation

by

Markus Sovala
Lic.Soc.Sc.

Academic dissertation to be presented, by the permission of the Faculty of Social Sciences of the University of Helsinki, for public examination, in Auditorium XII, Unioninkatu 34, on January 8, 1999, at 12 noon.

Helsinki 1998
ISBN:

951-45-8703-0 (PDF version)
Acknowledgements

Writing this book has been a long process. The whole staff of the Department of Economics at the University of Helsinki deserve my gratitude for their warm support during this work. I would especially like to thank Professor Seppo Honkapohja’s unofficial seminar group that consisted of himself, Professor Bengt Holmström during his sabbatical in Helsinki, Professor Matti Pohjola from the Helsinki School of Economics, Professor Mikko Puhakka from the University of Lapland, Dr. Pasi Holm, Dr. Jaakko Kian- der, Dr. Klaus Kultti, Dr. Urho Lempinen and Dr. Tuomas Saarenheimo. The main part of this study was written during 1990-1994 when I worked in a research project led by Professor Seppo Honkapohja.

The Faculty of Social Sciences nominated Professors Erkki Koskela and Matti Pohjola as official examiners for the manuscript of this study in 1994. I am extremely thankful for their patience in the reviewing process. They did not only require important changes in the text, but in many cases also suggested how to make them; reviewing became largely supervising.

I have got valuable comments to various drafts of this dissertation also from outside the department. Seminars organised by The Finnish Society for Economic Research, KAVA (Finnish Postgraduate Programme in Economics), University of Jyväskylä, International Economic Association, European Economic Association and Yrjö Jahnsson Foundation have been very stimulating. Comments by Professors Steinar Holden, Pertti Haaparanta and Tor Eriksson, as well as by Dr. Tuire Santamäki and Jaana Kurjenoja have not only taught me the limits of my analysis, but also encouraged me to continue forward.

This dissertation has its background in Docent Jukka Pekkarinen’s and Dr. Juhana Vartiainen’s research project on the History of Economic Policy in Finland. They introduced me into the profession of economic research. While I was working as a research assistant on the project, they suggested me to study the role of wage determination in macro-economic stability. My Master’s and Licentiate Theses, written under the supervision of Professor Koskela and Docent Pekkarinen, were results of this research. The discussions with Risto Vaittinen about my research topics have also been very helpful.

The subject of the last essay of the dissertation, i.e. the study of the Finnish wage arbitration system was suggested to me by Professor Matti Pohjola. He was also the person who helped, together with Jukka Pekkarinen, me to get into contact with Professor Bob Rowthorn and Dr. Ramana Ramaswamy at the University of Cambridge. I learned a lot from them.
and writing the essays on insurances and childcare in this dissertation was started during my visit to the Department of Economics and Politics in Cambridge in 1992-93. Väinö Tanner Foundation, Yrjö Jahnsson Foundation and Finnish Culture Foundation deserve my gratitude for financing the visit.

I am grateful to the Department of Economics, University of Helsinki, for including my research in its Research Reports series. Yrjö Jahnsson Foundation and the University of Helsinki have kindly financed the publication.

I thank the Academy of Finland and the aforementioned foundations for generous support and excellent research conditions that have enabled me to combine decent family life and research. My wife Dr. Laura Assmuth has been very interested in my research and my children Aino and Eero Assmuth have helped me to keep my feet on the ground. The eagerness of my children to make fun of my research ideas has been very valuable in understanding the limits of social science. I cannot thank too much my parents Anni and Juhan Sovala and my grandparents Alma and the late Viljo Sovala as well as my mother-in-law Sirkka Assmuth for support during this study.

Dr. Tuomas Saarenheimo, Juha Eskelinen, Dr. Tapio Bergholm, Dr. Markku Ollikainen, Pekka Bellof and Jorma Sappinen have acted like friends should; they all have encouraged my work also in the dark moments.

I shifted from the University to the Labour Institute for Economic Research when I was finishing the first draft of this dissertation. It was a soft landing. The same can be said about my current employer, Economics Department of the Ministry of Finance. Dr. Martti Hetemäki and my fellow civil servants have been very encouraging and I would like to thank all of them as well as the staff of the Government Institute for Economic Research. My fellow functionaries in the DG II of European Commission, whom I learned to know during my visit to Brussels in 1997-98, deserve the same thanks.

Jussi Huopaniemi skilfully helped me with language problems, but I am the only person to blame for any remaining mistakes. The same disclaimer applies, of course, to the substance of the study.

Helsinki, 7 December 1998

Markus Sovala
Studies on wage formation
by Markus SOVALA
1999
University of Helsinki
SF (Finland)
169 pages
ISBN 951-45-8420-1

The dissertation consists of three essays. The first one uses a simple implicit contract model to explain why risk-neutral firms buy insurance policies. By buying a policy, a firm gives a credible promise to repair damages and continue production in case of an accident, possibly against its own ex post motives. The higher expected employment can be traded against a lower wage.

The theme of the second essay is wage distribution and its links to childcare. Childcare requires similar kind of effort and care that employers demand at work. In the essay, the model of Ramaswamy and Rowthorn (Economica 1991) is augmented with gender specific effort functions. It is shown that women with higher domestic duties are allocated into the firms with relatively low dependency of effort and low wages. The higher the effort production cost for women is relative to that of men, the higher the wage gap is. The availability of public childcare facilities, e.g. daycare centres, might have an effect on wage gaps.

The third essay studies the Finnish wage arbitration system and compares it with those used in North America. The first argument defended in the essay is that the Finnish system of wage arbitration can be interpreted as a form of arbitration. This is not self-evident because there is a right-to-strike. The second argument concerns the role of arbitration. I argue that if there is some uncertainty about the award, the Finnish arbitration system shares the good informational properties of strikes or lock-outs, but on the other hand, the loss in the case of a bargaining impasse is also smaller than in the case of arbitration.
# Contents

1 Institutions and wage determination .......................... 5
  1.1 Background of the study ..................................... 5
  1.2 First essay: Why do risk-neutral firms purchase insurance policies? ..................................................... 7
  1.3 Second essay: Childcare and wage distribution .......... 9
  1.4 Third essay: Finnish wage arbitration - theoretical comments 10

2 Why do risk-neutral firms purchase insurance policies? 13
  2.1 Introduction .................................................. 13
  2.1.1 Paradox of corporate insurance ............................ 13
  2.1.2 Insurance policies change incentives ...................... 14
  2.1.3 Outline of chapter ....................................... 17
  2.2 Model of insurance policy demand .......................... 18
  2.2.1 Basic model .............................................. 19
  2.2.2 Efficient contract ........................................ 21
  2.2.3 Consequences of strong individual rationality .......... 23
  2.2.4 Insurance against bankruptcy ............................ 25
  2.2.5 Optimal insurance demand ............................... 27
  2.3 Related studies: A brief review ............................. 30
  2.3.1 Insurance purchases by risk-averse firms ............... 30
  2.3.2 Insurance policy as part of a contract ................. 31
  2.3.3 Trade union models ..................................... 33
  2.3.4 Implicit contract theory and beyond .................... 34
  2.4 Discussion .................................................. 37

3 Childcare and wage distribution ................................. 41
  3.1 Introduction .................................................. 41
  3.1.1 What is special in childcare? ............................. 42
  3.1.2 Domestic effort versus effort at work .................... 43
  3.1.3 Determinants of female childcare load .................. 44
3.1.4 Public daycare ........................................ 45
3.1.5 Changing family patterns ............................... 47
3.1.6 Traditional gender studies ........................... 48
3.1.7 Exogenous or endogenous gender differences? ... 49
3.1.8 Content and organisation of the chapter ........... 50
3.2 Basic model .............................................. 51
   3.2.1 General structure of the model ..................... 51
   3.2.2 Demand for labour and effort .................... 54
   3.2.3 Preferences of workers ............................ 57
   3.2.4 Equilibrium with one gender and two firms ...... 58
3.3 Heterogeneous workers .................................. 61
   3.3.1 Equilibrium with two genders - discrete case ... 62
   3.3.2 Equilibrium with two genders - continuous case .. 63
   3.3.3 Varying family commitments ...................... 66
3.4 Modernisation of family roles ......................... 68
3.5 Concluding remarks .................................... 72
   3.5.1 Predictions of the model .......................... 72
   3.5.2 Does the theory fit with the facts? ............... 72
   3.5.3 Is there any scope for change? ................... 75

4 Finnish wage arbitration - theoretical comments 77
   4.1 Introduction .......................................... 77
   4.2 The Finnish arbitration system ....................... 79
      4.2.1 Historical background .......................... 79
      4.2.2 Wage setting and arbitration in Finland ....... 81
   4.3 Arbitration in North America ....................... 84
      4.3.1 Forms of arbitration ........................... 84
      4.3.2 Conventional vs. final-offer arbitration ....... 86
      4.3.3 Empirical analyses ............................. 88
   4.4 Towards understanding arbitration ................. 91
      4.4.1 What is the right wage level? - A micro-economic view 91
      4.4.2 Reasons for labour disputes ..................... 97
   4.5 Arbitration as an implementation issue ............ 102
      4.5.1 Mechanism design: basic concepts .............. 103
      4.5.2 Incomplete information ......................... 107
      4.5.3 Complete information .......................... 110
      4.5.4 Arbitration and implementation ............... 113
   4.6 A model of the Finnish wage arbitration system ... 115
      4.6.1 The economic model and notations .............. 117
      4.6.2 Solution of the game ........................... 118
4.7 The function of the Finnish arbitration system .......................... 124
   4.7.1 Background ......................................................... 124
   4.7.2 The model and the solution ...................................... 125
   4.7.3 Solution of the game in a wider perspective ................. 128
4.8 Concluding remarks ..................................................... 131
   4.8.1 The Finnish system of wage arbitration ....................... 131
   4.8.2 Feedback mechanisms in arbitration ......................... 132
   4.8.3 Is binding arbitration an alternative for Finland? .... 134

References ................................................................. 137

A Appendix to Chapter 2 .................................................. 151
B Appendix to Chapter 3 .................................................... 155
C Appendix to Chapter 4 .................................................... 165
Chapter 1

Institutions and wage determination

1.1 Background of the study

This dissertation consists of three theoretical essays. The first one, Chapter 2, uses a simple implicit contract model to explain why risk-neutral firms buy insurance policies. The theme of the second essay, Chapter 3, is wage distribution and its links to different forms of childcare. The third essay, Chapter 4, studies the Finnish wage arbitration system. The essays deal with institutional aspects of wage determination. The analyses are based on the assumption of utility maximising behaviour. In this sense, the study belongs to the new generation of theoretical labour economic studies.

Modern studies in labour economics, especially in Europe, analyse the market structure of the labour market. The consequences of existence of strong organisations are analysed. Also, informational asymmetries have an increasingly important role in theoretical research. Although the new generation of research has been mainly theoretically motivated, empirical research in the field has also changed its outlook.

Recent developments in labour economics can be interpreted in many ways. I am inclined to think that labour economics is going back to its roots in concrete and institutionally aware research. Moreover, it is again admitted that that the history matters. However, it is now going back to its roots equipped with better theoretical tools and a deeper understanding of the nature of market relations and economic interdependence than what has been available before.

The development in labour economics has been very similar to general
developments in economics. A historically-oriented research pattern disappeared almost completely before the Second World War. In Finland, this happened later. The historical school was replaced by neo-classical micro-thy and Keynesian macro-economics, at least in academic research. This neo-classical synthesis, as it is sometimes called, became dominant in the whole Western world. The belief in this research agenda had its culmination point in the economic crises of the 1970’s. The oil crisis raised new issues for research and academic economists became, perhaps, more divided; competition between different macro-economic theories has been quite fervent since then.

In the late 1970’s and in the 1980’s, a major theme in theoretical economics was to find ”micro-foundations for macro-economics”. Though the major anomalies emerged in the field of macro-economics, the consequences in economic theory are well visible also in micro-economics. ”The micro-foundations” had to answer questions such as why are prices rigid, why do wages not decline when there is unemployment, why do financial markets not stabilise the economy, etc. Attempts to answer these and other questions generated a completely new body of micro-economic research. The availability of new theoretical tools, game theory and informational economics had a parallel effect. As a consequence, the assumption of atomistic agents was replaced by strategic interaction and the assumption of perfect information was replaced by the study of asymmetric information. At the moment, the micro-foundations for macro-economics are so well established that sometimes it is difficult to see any difference between macro and micro-theory.

The attempt to find a theoretically solid base for macro-economics has also changed the status of labour economics. One of the few widely accepted ideas among the economists’ profession is that rigidity of wages or too high a level of wages is somehow related to the chronic unemployment problem of the Western hemisphere. In that sense labour economics is now in a strategic position in the field of economic research.

The main challenge for theoretical analysis has been the need to explain why labour markets do not clear. There are four well-developed directions of research, each of which hold some promise of contributing to the answers. The first is the so-called ”implicit contract” theory. It builds on the assumption that firms are able to insure the workers against uncertainty by holding real wages relatively stable. The classical references are Azariadis (1975), Baily (1974) and Gordon (1974). In the 1980’s, implicit contract theory lost, however, its status as an interesting (un)employment theory when it was realised that the models did not predict unemployment but ei-
ther full or over-employment, depending on the informational assumptions. However, it is still accepted that these models may be relevant for other topics.

The second direction of research, which falls under the heading of "trade union models", is based on the observation that trade unions, or employed workers in general, have some bargaining power that leads to higher real wages and lower employment than would be observed under perfect competition. References to classical studies can be found from reviews by Oswald (1985) and Pencavel (1985), see also Layard, Nickell and Jackman (1991).

The third strand in theoretical literature is labelled "efficiency wages". According to these theories, the quality of labour is assumed to be related to the real wage. Yellen (1984) and Katz (1986) provide many relevant references, see also Bulow and Summers (1986) for interesting applications.

The fourth deviation from the perfect competition assumption is to assume that the labour market is characterised by "search" behaviour (see Mortensen 1986 and Sargent 1987).

The analyses presented in this study seek not to challenge the status of the major theory approaches in modern labour market studies. On the contrary, the models presented here use the achievements of the established theories to explain and analyse phenomena that are not well understood. Specifically, the analyses of this study use building blocks from implicit contract studies, efficiency wage theories and trade union models.

1.2 First essay: Why do risk-neutral firms purchase insurance policies?

The first essay analyses the question why do firms purchase insurance policies. In previous literature, this is explained by risk-aversion of the owners and by various institutional findings, including tax-related issues (Main 1983, Meyers and Smith 1982). Moreover, it has been argued that firms offer insurance for their employees because they cannot buy insurance themselves. Explanations are relevant in many cases and valuable as such, but it has been difficult to find explanations that are consistent with conventional micro-theory.

Chapter 2 aims to satisfy the demand for micro-economic explanations for firm insurance policy demand by offering a theoretical model where risk-neutral agents buy insurance policies. The explanation is based on the notion that a policy is a kind of a financial contract. By buying a policy,
a firm gives a credible promise to repair damages and continue production in case of an accident, possibly against its own motives. By an insurance policy, negative profit, i.e. loss, can be avoided in some cases and hence production is more likely to continue also after an accident. From the workers’ point of view, buying an insurance policy is a way of changing incentives of the firm. This motive has been brought up before in economic literature (e.g. see Mayers and Smith 1982), but the present analysis is the first attempt to model this analytically.

The purchase of an insurance policy is a way of committing \textit{ex ante} to do something that the firm does not want to do \textit{ex post}, i.e. rebuild the factory or repair the machinery after an accident and insurance can be used as a commitment device (the classical reference is Kydland and Prescott, 1977).

In the analysis of Chapter 2, the firm and its employees are assumed to be risk-neutral, but the workers are assumed to have some firm-specific human capital and thus are dependent on the continuous operation of the firm. The reason to assume risk-neutrality is simply a need to show as clearly as possible that the role of the insurance policy is not to reallocate risks but to alter incentives. The analysis of the chapter shows if the firm can deviate from the production agreement in the case of a fire or some other major accident, i.e. if it can go bankrupt, it is willing to include an insurance policy in the wage contract even if it is risk-neutral.

A simplified version of the implicit contract model of Rosen (1985) is used as a starting point for the analysis. It is shown that if the firm is able to cease production after an accident, the efficient contract cannot be maintained. An insurance policy providing for the minimum capacity for production after an accident improves the expected profit level, but it is nevertheless possible that the minimum capacity for profitable production cannot be provided when productive capacity is too low after an accident, i.e. when the damages are too large.

In the analysis, it is assumed that wages do not depend on the available production capacity. However, one may imagine a situation where wages can be adjusted in such a way that production is always profitable. In this case there would be no reason to buy an insurance policy as workers provide for the required insurance to the firm. There are reasons for this kind of behaviour not to be very common and for wages to be generally quite rigid. However, the question of possibility of applying conditional wages to overcome the problems which may arise when firms are able to cease production raises an alternative interpretation for the firms’ insurance policies. The insurance policies can be useful in reducing the utility loss
caused by non-flexible wages.

The use of price options as well as foreign exchange futures and forward contracts have become more popular recently. The thoughts presented in this chapter are, perhaps, useful in understanding why firms, that should be risk-neutral as the owners can insure their incomes via portfolio diversification, proceed in this direction. Though the share-holders of firms may be risk-averse, other stake-holders may not be.

1.3 Second essay: Childcare and wage distribution

The aim of Chapter 3 is to offer a theoretical explanation for gender wage differentials and labour market stratification. The model developed is also used for policy analysis. The analysis of the chapter starts with the observation that women are more likely than men to be responsible for childcare. Childcare requires similar kind of effort and care that employers demand at work. Hence, commitment to work according to the employers' rules is more demanding for female than for male employees. However, this difficulty is not constant over time or across societies. It depends, \textit{inter alia}, on the availability, the quality, and the costs of childcare facilities.

The basic nature of childcare implies that it is quite difficult to combine with professions characterised by high effort requirements. Professions not requiring total concentration or commitment are more easily combinable with heavy childcare loads.

It is reasonable to assume that workers demand better hourly compensation if they are asked to work harder or more efficiently. Because domestic duties vary, compensation demands of workers in different occupations also vary. As a consequence, we may expect that individuals with strong family commitments are allocated to jobs where effort is less important. Moreover, if pay is positively related to effort, then these people are also paid less.

The main building block of the model used in the analysis is adopted from studies by Ramaswamy and Rowthorn (1991, 1993). The authors generalised a basic efficiency wage model to take into account that the importance of effort varies between firms. Their model is capable of producing a wage distribution; firms valuing effort highly pay high wages and firms with time-intensive production technology maximise their profits by paying lower wages.

In Section 3.2 of this study, the Ramaswamy-Rowthorn model is aug-
mented with gender specific effort functions. In Section 3.3, it is assumed that there is a continuum of firms and that the firms differ from each other according to how important effort, i.e. flexible and careful work, is for production. It is shown that women with higher domestic duties are allocated into the firms with relatively low dependency of effort. They are also paid less than male employees. This is the main argument of the chapter.

The second main conclusion of the chapter concerns the gender utility difference, i.e. the gender gap. The higher the effort production cost for women is relative to that of men, the higher the gap is.

Many services are not available at the market; in Chapter 3 it is argued that childcare is such a service. The family has an absolute advantage in childcare, but social customs and public services available define how costly it is in terms of effort losses at the labour market. A heavy childcare load makes it difficult to work flexibly and carefully in a paid job. Historically, women have had the main load of childcare. This is partly because of biological reasons and partly because of social practices. Hence, the availability of public childcare facilities, e.g. daycare centres, might have an effect on wage gaps. This is the third main conclusion of the chapter.

1.4 Third essay: Finnish wage arbitration - theoretical comments

The third essay deals with the Finnish wage arbitration system. As far as I know, it is the first study on the topic with roots in economic theory.

Arbitration is widely used in North America, especially in public sector wage disputes, and there is ample economic research on how arbitration works and how it should be interpreted. Hence, it is natural to start the analysis with an review of the studies about North American arbitration experiences. Studies by Ashenfelter (1987), Ashenfelter and Currie (1990), Bloom and Cavanagh (1987), Chelius and Dworking (1980), Crawford (1979, 1981), Currie and McConnell (1991), Farber (1980), Farber and Bazerman (1986, 1987 and 1989), Farber, Neale and Bazerman (1990) are, perhaps, the most essential sources of institutional information and theoretical concepts.

The first argument defended in the chapter is that the Finnish system of wage arbitration can be interpreted as a form of arbitration. This is not self-evident because the Finnish system deviates from those in use in the United States and Canada and the rules of the Finnish system are quite
vague. However, Finnish labour law requires that all strikes and lock-outs should be announced to the State Arbitrator before the work stoppage is due. Moreover, participation in the arbitration is obligatory. Rules are well honoured and it can be argued that the whole system is implicitly designed, or at least accepted, by the labour market parties. In the essay, it is shown that if the award by the arbitrator is known beforehand and if it is preferred by both parties to the option of disputing, the arbitrator can effectively set the wage.

The second main argument of the chapter concerns the role of arbitration. It is inspired by a review article by Kennan and Wilson (1993), which is very critical of arbitration, as the authors emphasise the positive, information revealing role of work stoppages. The argument is very stimulating as it diverges from the mainstream thinking of arbitration studies. In Section 4.7, I argue that if there is some uncertainty about the award, the Finnish arbitration system shares the good properties of strikes or lock-outs. Uncertain outcomes of arbitration are not as effective an incentive to reach an agreement as work stoppages are, but on the other hand, the loss in the case of a bargaining impasse is also smaller than in the case of arbitration.

Although the functioning and the role of the Finnish arbitration system can be interpreted and understood in terms of abstract game theory, its structure and exact operation are quite unclear. There are several different phases in the arbitration and the timing of events varies from one year to another. Arbitration is in many senses based on traditions and implicit rules, as written procedural rules are scarce. Hence, theoretical arguments presented in this study must be based on the stylised facts of negotiation experiences and on strong simplifying assumptions. For example, The State Arbitrator has very little direct power in enforcing the bargaining code. Even the main aspect of the rules, namely the obligation to participate in arbitration, is very difficult to enforce as the parties may simply feign participation. However, parties seem to honour this and other written and unwritten rules quite well; this is also assumed in the theoretical analyses of the chapter.

The difficulty of applying formal game theory in modelling the Finnish arbitration system directed my analysis towards the third main conclusion of the chapter. It is argued that the Finnish arbitration system can be seen as a product and a combination of two different targets, or principles. First, the bargaining outcome should be effective and reflect the economic fundamentals as well as possible. The characteristics of good allocation rules are discussed in Section 4.4. On the other hand, the arbitration process can-
not be very complicated and should be robust in respect to opportunistic behaviour of the negotiating parties. These objectives are, however, contradictory. From abstract implementation theory, reviewed in Section 4.5, it is known that using simple strategies, only simple allocation rules can be implemented. More sophisticated allocation rules, for example the cooperative Nash solution, can be implemented only with dynamic games with strict rules for behaviour. Moreover, perfect information is required. But, as real negotiation situations are characterised by imperfect information and the negotiators have an incentive to signal more power than they really possess, for example by breaking the rules of game, allocations with fine theoretical properties are not normally implementable. Thus, the Finnish arbitration system is one particular solution to the trade-off between the targets, i.e. between the need to find a simple system and the need to produce allocations with desirable properties.
Chapter 2

Why do risk-neutral firms purchase insurance policies?

2.1 Introduction

2.1.1 Paradox of corporate insurance

Firms purchase insurance policies\(^1\). The policies vary, but there are some general features. In Finland, almost all firms have a fire insurance policy. The practice is so common\(^2\) that there is a general misbelief that fire insurance is obligatory. Insuring machinery or production as such is less common, but is everyday life in firms of different size, in different industries and with different market and ownership structures.

The situation may look paradoxical if one looks at it from the economist’s point of view. The basic reason to purchase an insurance policy is to reduce risks, but firms are normally seen in economics as risk-neutral entities. Though the owners may be risk-averse, the firm shouldn’t be. The owners’ risk-aversion apparently provides no incentive for the purchase of an insurance policy, since stockholders and other investors with access to capital markets can eliminate insurable risks through diversification of portfolios.

The reality on one hand and basic intuition of economists on the other

\(^1\)Business insurance accounted for 54 percent of direct property and liability insurance premiums written in the United States in 1978 (Mayers and Smith, 1982). Main (1983) cites a report by *Fortune*: ”. . . 86 percent of the top 500 industrial corporations purchased insurance against at least 75 percent of their perceived total loss exposure.”

\(^2\)In 1991, corporate fire and combined insurance (normally also fires are covered by combined policies) premiums totalled FIM 576 million (Järvikare 1996). See also Suomi-nen (1994).
seem to clash. The situation looks even more puzzling if one notes that expected profits are, in fact, increasing in price variability\(^3\) in a basic model of the firm\(^4\). Indeed, it has been difficult to find an explanation for corporate insurance demand from micro-theory\(^5\).

In the literature of insurance demand, however, several attempts have been made to explain why companies buy insurance policies. Explanations have been based on different institutional findings and observations, see discussion in Section 2.3. Many of the findings are plausible and add remarkably to our understanding of the behaviour of firms. However, there is a demand for more general explanations that have roots in micro-theory.

### 2.1.2 Insurance policies change incentives

This chapter aims to satisfy the demand for micro-economic explanations for insurance demand by offering a theoretical model in which risk-neutral agents purchase insurance policies. The explanation is based on the notion that an insurance policy is a kind of a financial contract. This motive has been mentioned before in economic literature (e.g. see Mayers and Smith 1982), but the analysis at hand is the first attempt to model this analytically. Previous studies dealing with related ideas are reviewed in Subsection 2.3.2.

An *ex ante* purchase of an insurance policy changes the incentives that the firm is facing *ex post*. Without a policy, the firm has to decide whether to continue production and whether to repair the damage occurred on the basis of reparation costs and variable production costs. With an insurance policy, the repairing costs are, at least implicitly, already paid and the continuation of production can be judged on the basis of future variable production costs only. Hence, some of the bankruptcies caused by adverse

---

\(^3\)The uncertainty in the model of this chapter is not related to production price but to damage of the physical productive capital. However, capital enters into the profit function in a similar way as the price level in the model analysed in this chapter.

\(^4\)Assume a firm having the profit function \(pF(L) - wL\), where \(p\) is a stochastic production price and \(w\) is a constant wage parameter. The convexity of the profit function guarantees that the expected profit from production under uncertain price conditions is higher than the profit with production price equal to the expected price.

\(^5\)An insurance policy may also change the expected values of, let us say, production capacity of the firm. If that is the case and the policy is purchased just to obtain higher expected production capacity, an additional question rises. Why does the firm not increase the expected value of its production simply by buying more capital? We will be more precise later (see Subsection 2.2.4) but it is worthwhile to affirm already now that this alternative is excluded from the analysis.
production conditions are avoidable.

A simplified version of the implicit contract model by Rosen (1985) is used as a starting point for the analysis. The strategy of the analysis is to show that an insurance purchase may be profitable even if all conventional motives are excluded. Though the formal analysis of the chapter is based on the assumptions of risk-neutrality of the agents, the purpose is not to claim that firms or workers are in fact risk-neutral; this topic is beyond the scope of this study. The reason to assume risk-neutrality is simply a need to show that the role of the insurance policy is not to reallocate risks but alter incentives.

In addition to the risk-neutrality of all agents, there are other assumptions that are intended to exclude all "trivial" motives to purchase an insurance policy. For example, the insurance company is assumed to not be superior in processing claims and reparation costs are assumed to be equal for everybody. Hence, there is no special reason to rely on an insurance company's help. Moreover, reparations of damages are assumed to be non-profitable in a conventional case, i.e. the price of new capital exceeds its marginal productivity. This extreme assumption is made to highlight the contractual motive for an insurance purchase. Without the assumption, it would turn out that the reparations are sometimes carried out by the firm, but sometimes only if it has an insurance policy. This would be an unnecessary complication.6

Though several possible reasons for insurance purchases are excluded, the basic setting of Rosen (1985), as well as the analysis presented here, may still exaggerate the motive to insure. It is based on the symmetric information assumption and thus the limitations caused by moral hazard and adverse selection problems are bypassed here.

It will be shown in Subsection 2.2.5 that under some feasible conditions, the firm is willing to include an insurance policy in the wage agreement. If the firm is able to cease production when actual profit is negative7, a large accident or a devastating fire may lead to factory closure and to a lay-off of workers. By an insurance policy, some cases of negative profit can be avoided and hence production is more likely to continue.

---

6The assumption that the price of capital exceeds its marginal productivity may be relevant for the firm with old production technology but reparation of modern machinery is more likely to be profitable as such. Hence, one may expect that contractual motives to buy insurance policies are more important in the case of outdated technology than in other cases.

7This possibility creates a strong individual rationality constraint (SIRC) for the contract, see Subsection 2.2.3.
From the workers' point of view, buying an insurance policy is a way of altering incentives of the firm. It can be thus argued that fire insurance policies can be seen as insurance against bankruptcies. The same mechanism can be analysed also from the firm's point of view. For example, the firm may purchase a fire insurance policy just to be able to make a credible promise to build the factory again after a possible fire. This kind of promise could be traded against a lower wage or a more profitable contract with a subcontractor. The trade union or the subcontractor would have reason to believe the promises only if the firm has an incentive to keep its word. Without an insurance policy, a promise to continue production after a fire or any other major damage would be without value if the total cost of production exceeds the value of production. The role of the insurance company would then be to guarantee that the firm honours its promise not to stop production after an accident. The insurance company could hence be seen as a third party which is used by the firm to reduce its future alternatives.

A purchase of an insurance policy is thus a way to commit *ex ante* to do something that the firm does not want to do *ex post*, i.e. rebuild the factory or repair the machinery and the insurance policy can be used as a commitment device (the classical reference is Kydland and Prescott, 1977). Tying one's own hands is known to be advantageous in some strategic situations, see references in Persson and Tabellini (1994).

It should be noted that the promise with guarantees is valuable only if the firm does not automatically repair the damages. If the reparations are profitable as such, there is no reason to purchase an insurance policy. Altering the incentives the firm is facing may be valuable only if the workers or other "stake-holders" value increased job security highly. High valuation may arise when repairs are profitable *ex ante*, i.e. if expected costs of the repairs are lower than the gain from higher expected employment, but non-profitable *ex post* from the firm's point of view. In this kind of a situation an insurance policy may solve the time consistency problem.

One cornerstone of the analysis deserves comment. It was assumed that the insurance purchase is included in the wage contract. There is, however, no evidence that such contracts exist. Hence, the contracts, if they exist, must be of implicit nature. Implicit contract theories (see Subsection 2.3.4) have been criticised from similar bases; written "implicit contracts" do not exist. The situation concerning insurance policies is, nevertheless, very different. Formal inclusion of a policy into a wage contract is not necessary for two reasons. First, there is no reason to include it in an explicit contract for possible use in court, as the insurance company guarantees the deal. Second, because insurance contracts are usually long-term contracts, the
insurance decision of the firm is already known before the wage negotiation starts.

It is important to note that the value of job security may result from other factors than risk-aversion. A credible promise of a job with no risk of redundancy may be valuable if there is friction in the labour market or if the workers have firm-specific skills. Then the risk of losing a job decreases expected wage income. This has a negative effect on the expected utility of the workers even if they are risk-neutral.\footnote{The following example may provide an explanation why insurance policies are important and in fact are very common. Consider a case, in which you are asked to move to a small town with only one factory to take up a job in that factory. The factory is outdated and your work will require firm specific skills, which you only can learn by doing and which are of no value outside the firm. How would you react if you were told that the firm does not have a fire insurance policy?}

The insurance purchase can, hence, be seen as a solution to the problem of deviation between the private value and the social value of production. Production after a fire may be unprofitable for the firm because of high repairing costs. At the same time, the workers may value production highly because an alternative income may be low. If both components of social value are included, the value of production may exceed the sum of repairing and production costs.

\subsection*{2.1.3 Outline of chapter}

The basic argument behind the analysis is explained and the formal model is presented in Section 2.2. It will be shown that under some quite non-restrictive assumptions, an insurance policy is always included in a wage agreement.

After the analysis of contractual factors in the insurance purchases, previous insurance studies are reviewed in Subsections 2.3.1 and 2.3.2. The results of Section 2.2 are compared with the outcomes of the trade union and the implicit contract models in Subsections 2.3.3 and 2.3.4. Section 2.4 concludes the analysis.
2.2 Model of insurance policy demand

There are several ways to model insurance policy purchases by risk-neutral firms. The basic implicit contract model\(^9\) of Røsen (1985) is adopted here as a starting point. The model is originally from Azariades (1975) and simple enough to produce tractable results. However, in the analysis at hand, profit of the firm, not utility of the workers, is maximised. Moreover, Røsen’s model is simplified by assuming linear production technology and a uniform distribution for the productivity parameter.

In Subsection 2.2.2, the model is solved and the properties of optimal contract are studied. The efficient solution of the model requires only that the workers are willing to sign a contract with the firm, i.e. that the expected utility exceeds or is equal to that what is available elsewhere in the economy. Actual profit may be negative if the realised productivity is very low, perhaps after an accident or a fire at the plant.

In Subsection 2.2.3, the analysis deviates from the implicit contract model and deals with a more realistic case. It is required that the profit of the firm should be zero or positive with all realisations of productivity. If not, the firm ceases production and faces bankruptcy; in this case all workers will be made redundant. This condition is called the *strong individual rationality constraint*. It will be shown that the efficient contract cannot be maintained if the constraint is honoured. Production is lost if the damages are large, i.e. if the productive capital is low. The minimum level of productive capital for profitable production depends on production costs, i.e. compensations for the workers. Hence, the constraint is analysed in the form that defines the minimum amount capital required for profitable production as a function of monetary compensation to the workers.

A possibility to repair the damages is introduced in Subsection 2.2.4. This would imply that also the expected value of productivity can be altered. However, it will be assumed that the reparations are in general non-profitable, i.e. the price of new capital exceeds the marginal product of the capital. The first conclusion is, of course, that there is no reason to buy an insurance policy if there are no contractual inefficiencies. The situation is, however, different when the strong individual rationality requirement constrains the terms of the contract. In Subsection 2.2.5, it will be shown that an insurance policy purchase can be used to decrease the utility loss resulting from contractual inefficiency.

\(^9\)The implicit contract models are discussed with some more depth in Subsection 2.3.4.
2.2.1 Basic model

In the model of Rosen (1985), the firm makes a contract with a group of identical workers. Output depends on the amount of utilised labour and the stochastic disturbance \( \theta \). The firm is assumed to pay a salary to its employed workers and an unemployment benefit to its unemployed workers\(^{10}\). The optimal labour contract is derived by maximising the utility of a representative worker subject to the reservation profit of the firm. The contract defines the wage, the unemployment benefit and the employment rule for every realisation of the stochastic \( \theta \).

Our analysis deviates here from that of Rosen. First, in the analysis at hand, the profit of the firm is maximised, not the utility of the workers. It is more natural to maximise profit when one tries to explain behaviour of the firm. Second, we take into account the very fact of the real world that wages are not very often indexed to production conditions. There is a large theoretical and empirical literature on nominal rigidities of wages, but these studies are not reviewed here. For our purposes, it is enough to note that conditional wage schemes are rare and that this is partly due to informational problems\(^{11}\). Moreover, it will turn out that risk-neutrality of the firm implies constant wage in the basic implicit contract model.

Rosen adopted the same simplifying assumption about the preferences of the workers as Azariadis did in his seminal paper (1975). The utility of a worker is 
\[
    u = U(C + mL),
\]
where \( C \) is consumption and \( L \) is the fraction of time devoted to leisure and \( m \) is a constant parameter. \( L \) is normalised such that \( 0 \leq L \leq 1 \). This utility function has linear indifferent curves and \( C \) and \( L \) are perfect substitutes. The parameter \( m \) measures the marginal product of producing non-market goods or equivalently the monetary value of leisure. Hence, \( m \) is the reservation price of time supplied to work.

To exclude conventional motives to purchase an insurance policy, the following assumption is essential.

**Assumption 2.1** The workers are risk-neutral and the utility function \( U(\cdot) \) is linear such that
\[
    u = C + mL.
\]

\(^{10}\)Alternatively, the unemployed workers may receive unemployment benefit from the government. Analysis of this situation would require defining how the benefit is financed and is hence too complicated for our purposes.

\(^{11}\)One may wonder whether the informational problems cause the same problems for the employment function \( \rho(\theta) \). This is not the case, as the firm has an incentive to follow the employment rule because it is, by construction, a condition for profit maximisation.
Following Rosen (1985), we require that the worker has to work full-time or not at all, that is $L = 1$ or $L = 0$. The firm has a constant number of workers and uses a fraction $p$ of them in production. The firm sets $p$, which in general is a function of $\theta$, i.e. $p = \rho(\theta)$. Thus $0 \leq \rho(\theta) \leq 1$.\footnote{Rosen assumes that number of workers is $n\rho(\theta)$. Here we make an implicit assumption that $n \equiv 1$.}

The contract defines the compensation of the workers. We denote wage by $C_1$ and unemployment benefit by $C_2$. As noted already, the compensations are constants, i.e. they do not depend on the available production capacity.

**Assumption 2.2** Utilities of the workers are constant across the employment status, e.g.

$$C_1 = m + C_2.$$  

This condition is natural. If it does not hold, there is no equilibrium at the internal labour market of the firm as the unemployed (employed) would start to compete for jobs (leisure) by offering contracts with wage $C_1 - \xi$ ($C_2 - \xi$) if $C_1 > C_2 + m$ (if $C_1 < C_2 + m$), where $\xi$ is a small positive number. From this basis, it is natural to assume that every worker has the same probability to be employed, i.e. that they are allocated randomly.\footnote{It is interesting to compare this assumption with the result derived from basic implicit contract theory. Risk-aversion of the workers guarantee in the basic implicit contract model that employed and unemployed workers enjoy the same level of income for any given value of $\theta$, i.e. $C_1(\theta) = m + C_2(\theta)$. In addition, if the firm is risk-neutral and thus able to provide full insurance coverage for the workers against all variations in incomes, optimal wage contract is of the form $C_1 = m + C_2$.}

The damages are modelled by a variable production capacity in the production function. If one denotes full production capacity with $C$ and stochastic damages with $D$, then the capacity available for production can be expressed as $\theta = C - D$. In what follows, the analysis is carried out by using variable $\theta$.

**Assumption 2.3** Production function of the firm is of the form

$$x = \theta \rho(\theta),$$  

where $\theta$ is the stochastically determined production capacity and $\rho(\theta)$ labour input.

For notational ease, it is assumed that $\theta$ is restricted such that $0 \leq \theta \leq 1$. Realisation of $\theta = 1$ implies that there is no damage to the production plant.
and low values of $\theta$ correspond to large damages. Later, we will analyse more carefully how $\theta$ depends on the existence of insurance policies. Before that, it is enough to make the following assumption\(^{14}\).

**Assumption 2.4** $\theta$ is a continuous random variable taking values between 0 and 1 with density function $g(\theta) = 1$, i.e. $\theta$ is uniformly distributed between 0 and 1.

Because employed workers do not have any leisure, their utility is simply $C_1$. 1$-\rho(\theta)$ unemployed workers enjoy leisure and receive an unemployment benefit $C_2$. Hence, the expected utility of a worker is

$$Eu = \int_0^1 [C_1 \rho(\theta) + (C_2 + m) (1 - \rho(\theta))] d\theta.  \tag{2.1}$$

Assumption 2.2 implies that

$$Eu = C_2 + m.$$

Using function $\rho(\theta)$ and equality $C_1 = C_2 + m$, profit of the firm can be expressed as a function of $\theta$

$$\pi(\theta) = \theta \rho(\theta) - \rho(\theta) C_1 - (1 - \rho(\theta)) C_2$$

$$= \theta \rho(\theta) - \rho(\theta) m - C_2 \tag{2.2}$$

$$= y(\theta) - C_2,$$

where $y(\theta)$ is the social value of production. Expected profit is

$$E\pi = \int_0^1 y(\theta) d\theta - C_2. \tag{2.3}$$

### 2.2.2 Efficient contract

An efficient contract maximises the expected profit of the firm with respect to $C_2$ and the employment function $\rho(\theta)$, given the alternative utility of the workers. Because the utility of leisure is $m$, it is natural to assume that the reservation utility $M$ exceeds the value of leisure.

\(^{14}\)Rosen assumes only that $\theta$ is distributed with a known distribution function $G(\theta)$ and a density function $G'(\theta) = g(\theta)$. 

Assumption 2.5 $M > m$, i.e. the value of workers’ outside option is higher than the value of leisure.

If Assumption 2.5 does not hold, production would not take place as the workers prefer to stay at home.

The contract is obtained by solving the following problem.

$$
\max_{C_2, \rho(\theta)} \int_0^1 \pi(\theta) \, d\theta \quad \text{s.t.} \quad Eu = m + C_2 \geq M
$$

(2.4)

Noting that $\pi(\theta) = \theta \rho(\theta) - \rho(\theta)m - C_2$ allows us to conclude that optimal labour demand is

$$
\rho^*(\theta) = \begin{cases} 
0 & \text{if } \theta < m \\
1 & \text{if } \theta \geq m
\end{cases}
$$

(2.5)

and that optimal $C_2^*$ is the lowest number satisfying $C_2 \geq M - m$, i.e. $C_2^* = M - m$.

The firm pays compensation $C_2^* = M - m$ to the workers in any case and an additional compensation $m$ if there is production. Thus profit

$$
\pi(\theta) = \begin{cases} 
-C_2^* & \text{if } \theta < m \\
\theta - m - C_2^* & \text{if } \theta \geq m
\end{cases}
$$

(2.6)

is negative if $\theta < M$. Straightforward calculation shows that expected profit is

$$
\int_0^1 \pi(\theta) \, d\theta = \frac{1}{2} + \frac{m^2}{2} - M.
$$

(2.7)

Hence it is clear that the firm has positive expected profit only if $M$ is small enough.

It is important to note that in the basic case analysed here there is no scope for an insurance policy purchase as the firm has no incentive to reduce the variability of $\theta$. For example, if an insurance policy guaranteeing that production capacity $\bar{\theta} = E\theta = 1/2$ is available in all circumstances, expected profit would be

$$
\frac{1}{2} - m - C_2 = \frac{1}{2} - m - M + m = \frac{1}{2} - M,
$$

(2.8)

---

15 Because $C_1 = m + C_2$, the expected utility of the workers is $m(m + C_2) + (1 - m)C_1 = m + C_2$, where $1 - m$ is the probability that production takes place.

16 If production takes place, the workers get $C_1^* = m + C_2^* = M$. 
2.2. Model of Insurance Policy Demand

provided that $m < 1/2$. This certain profit is smaller than the expected profit under stochastic productivity, because the opportunity to avoid production under unfavourable production conditions is lost. Moreover, the workers do not gain if this kind of insurance policy is bought in this case as the expected utility remains at the level $M$.

2.2.3 Consequences of strong individual rationality

Until now it has been assumed that an efficient contract can be agreed upon and implemented without any problems. However, in the real world there are informational and contractual inefficiencies that hinder the first-best solutions. In what follows, one particularly important contractual imperfection is analysed more carefully.

It is reasonable to assume that the firm may stop producing if production is causing a loss; i.e. that bankruptcy is possible. In this situation, the firm would not pay the wage $C_1$ to the employed nor the compensation $C_2$ to the unemployed workers. Hence, it is required that the firm should produce non-negative profit for every realisation of $\theta$ (i.e. $\pi(\theta) \geq 0$), otherwise there would be no production. This constraint is called the strong individual rationality constraint, SIRC.$^{17}$

Consider first the basic case without a possibility to buy an insurance policy. The contract can be obtained by solving the following problem.

$$\max_{C_2, \theta^c, \rho(\theta)} \int_{\theta^c}^{1} \pi(\theta) \, d\theta$$

s.t. $m + C_2 \int_{\theta^c}^{1} \, d\theta \geq M$ \hspace{1cm} (2.9)

$$\pi(\theta^c) = 0,$$

The problem is the same as (2.4) on page 22, but there is an additional restriction $\pi(\theta) \geq 0$. If the restriction is not fulfilled, there is no production.

$^{17}$One may wonder whether the constraint can be bypassed via capital markets. It should, however, be noted that this would not be a solution. The firm is not producing when $\pi(\theta) < 0$ because it maximises profit and, by assumption, it can always obtain zero profit by not producing. Hence, even if there is financing available for negative profits, there is no reason produce if $\pi(\theta) < 0$. Moreover, banks do not have an incentive to finance losses. The insurance companies are prepared to do this only because they receive insurance premiums before production takes place.
Because \( \pi (\theta) \) is increasing in \( \theta \), the restriction can be written in a more tractable form \( \pi (\theta^C) = 0 \). The probability that production is carried out is thus

\[
\int_{\theta^C}^{1} d\theta = 1 - \theta^C, \tag{2.10}
\]

where \( \theta^C \) is the threshold for profitable production.

The optimal contract has the following properties. The profit function \( \pi (\theta) = \theta \rho (\theta) - \rho (\theta) (m - C_2) \) can have positive value only if there is production, i.e. only if \( \rho (\theta) > 0 \). Moreover, because of linearity of production technology, \( \rho (\theta) \) is either zero or one. Thus \( \theta \rho (\theta) - \rho (\theta) (m - C_2) = \theta - m - C_2 \) must be positive and the threshold for the profitable production is then \( \theta^{C^{**}} = m + C_2 \) and optimal labour demand is

\[
\rho^{**} (\theta) = \begin{cases} 
0 & \text{if } \theta < m + C_2 \\
1 & \text{if } \theta \geq m + C_2.
\end{cases} \tag{2.11}
\]

When \( \rho^{**} (\theta) \) and \( \rho^{*} (\theta) \) are compared, one observes that SIRC causes a loss of production when \( \theta \in [m, m + C_2) \). The corresponding profit function is

\[
\pi (\theta) = \begin{cases} 
0 & \text{if } \theta < m + C_2 \\
\theta - m - C_2^{**} & \text{if } \theta \geq m + C_2.
\end{cases} \tag{2.12}
\]

Because the probability of profitable production is \( 1 - \theta^{C^{**}} = 1 - m - C_2 \), the optimal \( C_2^{**} \) can be obtained by noting that it is the lowest number satisfying \( C_2 (1 - m - C_2) \geq M - m \).\(^1\) If \( M \) is very large, it may be possible that such a number does not exist and no contract is signed. Because \( C_2 (1 - m - C_2) \) reaches its maximal value when \( C_2 = (1 - m) / 2 \), production can be maintained only if

\[
M \leq \frac{(1 - m)^2}{4} + m. \tag{2.13}
\]

If an agreement is reachable, the expected profit is

\[
\int_{m + C_2}^{1} \pi (\theta) d\theta = \frac{(1 - m - C_2^{**})^2}{2}. \tag{2.14}
\]

---

\(^{18}\) See Equation (2.5) at page 22.

\(^{19}\) Note that utility of the unemployed workers is \( m \) and the utility of employed workers is \( C_1 = m + C_2 \) and that the corresponding probabilities are \( m + C_2 \) and \( 1 - m - C_2 \). Hence, it is required that \( m (m + C_2) + C_1 (1 - m - C_2) \geq M \) or equivalently \( C_2 (1 - m - C_2) \geq M - m \).
Proposition 2.1 (a) The expected profit under SIRC is lower and (b) the wage higher than in the efficient case.

Proof See Appendix A.

2.2.4 Insurance against bankruptcy

In Subsection 2.2.2, it was noted that an insurance policies which reduces variability of productivity does not increase utility of the contracting partners. A policy may, however, also change the expected value of production capacity. For example, an insurance policy may guarantee full production capacity for the firm in all circumstances. This kind of insurance would be profitable if the insurance premium is low compared to the expected increase of production capacity.

The lower bound for the insurance premium is expected reparation costs. If the insurance company is offering actuarially fair policies, i.e. if the insurance market is competitive, the insurance premiums coincide the expected reparation costs. In what follows, this is assumed.

Let us start with a general definition of the insurance policy $\beta(\theta)$. The insurance company receives an insurance premium and promises to repair some part of the losses caused by an accident or a fire. From the firm’s point of view, this implies higher operating capital and higher productivity than would be available if there is no insurance. The available production capacity $\theta + \beta(\theta)$ is determined by realised $\theta$ drawn from the distribution $g(\theta)$ and by the insurance policy $\beta(\theta)$. The production function (see Assumption 2.3 on page 20) becomes

$$ x = \theta \rho(\theta) \equiv \left[ \theta + \beta(\theta) \right] \rho(\theta + \beta(\theta)), \quad (2.15) $$

where $\theta$ is shorthand notion for $\theta + \beta(\theta)$ and function $\beta(\theta)$ describes the scale of reparations, i.e. the insurance policy. The function is bound such that $0 \leq \beta(\theta) \leq 1 - \theta$ for every $\theta$. For example, if the insurance company is committed to repairing all damages, then $\beta(\theta) = 1 - \theta$ and the firm always operates with full capital stock 1. If there is no insurance, $\beta(\theta) \equiv 0$ for every $\theta$.

The expected cost of reparations is

$$ K = q \int_{0}^{1} \beta(\theta) d\theta, \quad (2.16) $$
where $q$ is the price of investment goods. Our assumption on actuarially fair policies implies that this is also the insurance premium.

Substituting (2.15) into (2.3) from page 21 and taking (2.16) into account gives

$$E\pi - K = \int_0^1 [\theta \rho (\theta) - \rho (\theta) m - q \beta (\theta)] d\theta - C_2. \quad (2.17)$$

If $C_2$ is taken as given\textsuperscript{20}, the partial derivative of the expected net profit (after the insurance costs) with respect to $\beta (\theta)$ is\textsuperscript{21}

$$\rho (\theta) - q \quad (2.18)$$

If the price of capital is high enough, there is no reason to include insurance into the contract if $C_2$ is given. Because $0 \leq \rho (\theta) \leq 1$, condition $q > 1$ would guarantee that the firm does not want to buy an insurance policy if there are no contractual inefficiencies. To exclude all trivial motives to buy an insurance policy, it is assumed that the following holds.

**Assumption 2.6** The price of capital exceeds its marginal productivity, i.e. $q > 1$.

If this assumption is not made, it may turn out that the firm would be interested in buying an insurance just to increase its expected production capacity as the price of capital is so low that investments are profitable as such. Then the demand for insurance policies would be higher than with the assumption. The assumption is thus intended to produce a case where insurance policies only have a contractual role.

Consider the following insurance policy.

$$\hat{\beta} (\theta) = \begin{cases} \theta^C - \theta & \text{if } \theta \in [\theta^C - \varepsilon, \theta^C] \\ 0 & \text{otherwise} \end{cases} \quad (2.19)$$

\textsuperscript{20}The analysis of Subsection 2.2.5 is based on the observation that $C_2$ is not fixed when an insurance is introduced.

\textsuperscript{21}This method may be unfamiliar. Think of the integral in (2.17) as the limit of a sum across a large number $T$ of discrete possible realisations of $\theta$. For example, the integral $A = \int_0^1 F (\theta) g (\theta)$ can be written in the form $A = \sum_{i=1}^T F (\theta_i) p_i$, where $p_i$ is probability of the event $\theta = \theta_i$. Then $T$ partial derivatives $\frac{\partial A}{\partial \beta (\theta)} = F' (\theta_i) g (\theta)$ can be written compactly $\frac{\partial A}{\partial \beta (\theta)} = F' (\beta (\theta)) g (\theta)$. See footnote 3 on page 1151 in Rosen (1985) for further details.
This policy provides the minimum capacity for profitable production if the realisation of $\theta$ is not too low. The idea is that the productive capacity is replaced after a damage or a fire up to a limit $\theta^C$, i.e. up to the limit of profitable production. However, reparations are not carried out if they require an increase in capital larger than $\varepsilon$. As a result, the actual productive capacity in use is $\theta$ if realised $\theta$ is smaller than $\theta^C - \varepsilon$ or larger than $\theta^C$. If $\theta$ is between $\theta^C - \varepsilon$ and $\theta^C$, the available capacity is $\theta^C$.

Price of the policy $\hat{\beta}(\theta)$ is

$$K = q \int_{\theta^C - \varepsilon}^{\theta^C} \left[\theta^C - \theta\right] d\theta = \frac{q\varepsilon^2}{2} \quad (2.20)$$

where $q$ is price of investment goods. Production can be maintained by an insurance policy $\hat{\beta}(\theta)$ if the realisation of production capacity is $\theta^C - \varepsilon$ or higher.

### 2.2.5 Optimal insurance demand

The tools developed above can be used to derive the main conclusions of this chapter. If an insurance policy of type $(2.19)$ is available, production can be maintained when $\theta \geq \theta^C - \varepsilon = m + C_2 - \varepsilon^2$. The linearity of the production function guarantees that if there is production, $\rho(\theta) = 1$ and all workers are employed.

Though production can be maintained, it is not necessarily profitable as $\pi(\theta) = 0$ when $\theta \in [\theta^C - \varepsilon, \theta^C]$. If insurance costs are neglected, profitable production requires that $\theta \geq \theta^C = m + C_2$. Expected profit without insurance costs is then

$$\int_{m+C_2}^{1} \pi(\theta) d\theta = \frac{(1 - m - C_2)^2}{2} \quad (2.21)$$

and with insurance costs

$$\frac{(1 - m - C_2)^2}{2} - \frac{q\varepsilon^2}{2}. \quad (2.22)$$

Because the workers’ participation constraint is $C_2 (1 - m - C_2 + \varepsilon) \geq M - m$, where $(1 - m - C_2 + \varepsilon)$ is the probability that there is production

\footnote{Note that $C_2$ need not be the same as $C_2^\pi$ obtained in Subsection 2.2.3.}
and the compensations are paid\textsuperscript{23}, the optimal contract can be obtained by solving the following problem\textsuperscript{24}.

\[
\max_{C_2, \varepsilon} \quad \frac{(1 - m - C_2)^2}{2} - \frac{q\varepsilon^2}{2} \\
\text{s.t.} \quad C_2 (1 - m - C_2 + \varepsilon) - M + m \geq 0
\]

(2.23)

The first-order conditions of the problem are

\[
\lambda = \frac{1 - m - C_2}{1 - m - 2C_2 + \varepsilon} \quad (2.24)
\]

for \(C_2\) and

\[
\lambda = \frac{q\varepsilon}{C_2} \quad (2.25)
\]

for \(\varepsilon\). The multiplier \(\lambda\) is related to the workers’ participation constraint. Because (2.24) is decreasing and (2.25) increasing in \(\varepsilon\), there are unique \(\varepsilon\) and \(\lambda\) solving the system of the first order conditions. The obtained

\[
\varepsilon^{***} = \frac{\lambda C_2}{q} \quad (2.26)
\]

determines \(C_2^{***}\) and hence also \(C_1^{***} = C_2^{***} + m\) uniquely.

**Proposition 2.2** The optimal contract has the following properties:

(a) The firm buys an insurance policy with \(0 < \varepsilon^{***} < C_2^{***}\).

(b) The optimal \(C_2^{***}\) is the smallest number satisfying

\[
C_2 (1 - m - C_2 + \varepsilon) \geq M - m.
\]

(c) Production takes place with full staff if \(\theta \geq m + C_2 - \varepsilon\).

(d) The firm is able to increase its expected profit by buying an insurance policy.

\textsuperscript{23}Unemployed workers’ utility is simply \(m\) because no unemployment compensations are paid as the firm is bankrupt and employed workers enjoy an income of \(C_1 = C_2 + m\). The participation constraint \(C_2 (1 - m - C_2 + \varepsilon) \geq M - m\) is obtained on the basis of these by straightforward calculation.

\textsuperscript{24}The problem reduces to that given in (2.9) if \(\varepsilon \equiv 0\).
Proof See Appendix A.

Inspection of the workers’ participation constraint reveals interesting properties. First, if reservation utility $M$ is very large, there is no way of reaching an agreement. Because $C_2 (1 - m - C_2 + \varepsilon)$ has its maximal value when $C_2 = (1 - m + \varepsilon) / 2$, an agreement can be reached only if

$$M \leq \frac{(1 - m + \varepsilon)^2}{4} + m. \tag{2.27}$$

Comparison with Equation (2.13) (see page 24) reveals that there are values of $M$ with which production is possible only if there is an insurance policy available.

Second, because the constraint is binding, one may set

$$C_2 (1 - m - C_2 + \varepsilon) - M + m = 0 \tag{2.28}$$

and calculate

$$\frac{dC_2}{d\varepsilon} = -\frac{C_2}{1 - m - 2C_2 + \varepsilon}. \tag{2.29}$$

Hence the introduction of an insurance would decrease the wage $C_2$ if $1 - m - 2C_2 + \varepsilon$ is positive. Then also the probability of profitable production increases as $1 - m - C_2 + \varepsilon$ is increasing in $\varepsilon$ and decreasing in $C_2$. One may note that $1 - m - 2C_2 + \varepsilon$ is indeed positive. If it were negative, $C_2 (1 - m - C_2 + \varepsilon)$ would be decreasing in $C_2$. Hence, we can conclude that $dC_2/d\varepsilon$ is negative if production is possible.

The above discussion justifies the following proposition.

Proposition 2.3 If SIRC binds, i.e. if there is a possibility for the firm to go bankrupt, the firm buys an insurance policy. As a consequence, the firm can increase the probability of production and pay smaller wages. In other words, $1 - m - C_2^{***} + \varepsilon > 1 - m - C_2^{**}$ and $C_2^{***} < C_2^{**}$.

Though the firm is able to reduce losses due to the strong individual rationality constraint by buying a policy, some of the losses still remain.

Proposition 2.4 An insurance purchase cannot eliminate all inefficiencies caused by a possibility of bankruptcy. With a policy,

(a) the wage level is higher,
(b) the probability of production is lower, and
(c) the expected profit is lower than in the efficient case.

Proof See Appendix A.
2.3 Related studies: A brief review

Before entering the conclusions, studies discussing the "traditional" as well as contractual motives of buying insurance policies should be reviewed. In addition to this, the results of Section 2.2 are compared with the outcomes of the trade union and the implicit contract models in Subsections 2.3.3-2.3.4.

2.3.1 Insurance purchases by risk-averse firms

The problem of firms' insurance demand is known in the literature and many authors have tried to explain why firms buy insurance policies\textsuperscript{25}. As Mayers and Smith (1982) pointed out, the task is fairly simple for the closely held firm or for the firm owned by one family. These firms are likely to purchase an insurance policy for the same reasons as individuals buy insurance policies. Hence, the theory of consumers' insurance demand is widely applicable if the firm is risk-averse. However, one should keep in the mind that other economic factors may override risk-aversion. For example, if profit can be very high in some cases, even risk-averse owners of the firm may prefer uncertain production conditions to certain ones.

Consumers' motives of purchasing insurance policies are analysed in several studies, for classical treatments see Mossin (1968), Ehrlich and Becker (1972) and Arrow (1974). The basic result of the analyses is that the less risk-averse contracting party bears more risk than the more risk-averse one. Insurance companies are capable of reducing aggregate risk by pooling numerous individual consumers' risks together and diversifying the remaining risk by wide ownership and pooling contracts. As a consequence, the insurance companies are often assumed to be risk-neutral. Risk-neutral insurance companies are capable of offering full-coverage insurance to consumers if the probabilities of accidents are independent of actions taken by the consumers or if the actions are observable and verifiable. The situation is, however, very different if the clients' actions are unobservable and affect the probabilities. A moral hazard problem arises and the full-coverage is optimal only if the cost of careful behaviour is very large or approaches zero. (Shavell 1979)

\textsuperscript{25} Articles are mainly published in the Journal of Risk and Insurance and Journal of Business. The only Finnish studies on the topic are theses by Järvikare (1994,1996).
Moreover, if the clients are of different types and the type of an individual is unobservable, an adverse selection problem may arise and limit the use of insurance contracts. If risk-aversion of the clients is relatively mild and the insurance pool contains a relatively wide range of risks, the insurance company is unable to offer full coverage to all consumers. If the "bad" risks are covered, the insurance becomes too expensive for "good" consumers. (Hirshleifer and Riley, 1992)

2.3.2 Insurance policy as part of a contract

The study at hand is not the first attempt to analyse insurance policies as contractual entities. According to the Modigliani-Miller Theorem, corporate financing policy is irrelevant if a firm’s investment policy is given and if there are no contracting costs or taxes. For the firm, an insurance purchase is just one element of its financing policy. Taking a loan or selling shares provide capital immediately and unconditionally, but insurance policies are similar to options or futures. These financial instruments provide capital for transactions or reinvestments with delay\(^ {26} \) and their usefulness is not known beforehand. Making reference to Modigliani and Miller, Mayers and Smith (1982) conclude that the firm’s insurance policy is of importance only if it is so because of taxes, because of the impact of an insurance policy on the firm’s investment decisions or on contracting costs.

Governments seem to favour the purchase of insurance policies. In addition to direct regulation, tax systems include incentives to buy policies (Main 1983, Mayers and Smith 1982). However, as Mayers and Smith (1981) argue, this does not explain the existence of insurances, but raises the next logical question: why does regulation favour this institutional form?

The relation between the model of this chapter and models of implicit contract theory is discussed in more depth in Subsection 2.3.4, but it is worthwhile to note that the basic form of implicit contract theory does not focus on contracting costs. If contracting is not costly, it is unimportant how the property rights are allocated. However, in the real world contracting is expensive. Because the capital stock of the firm limits the amount of risk that can be allocated to stockholders and bondholders, the insurance contract renders an efficient allocation of risk for the firm’s other claim holders. Mayers and Smith (1982) suggest that the higher the employees’, customers’, and suppliers’ fraction of the claims to the firm’s output, the higher the probability that the firm will purchase an insurance policy.

\(^ {26} \) Also retroactive insurance policies exist, see Smith and Witt (1985).
A study by Parry and Parry (1991) is an application of the above-mentioned reasoning. The authors study the purchase of an insurance policy by a risk-neutral firm\(^{27}\) for a risk-averse prospective employee. To hire an employee, especially a manager, the firm has to provide a compensation package consisting of a salary and an insurance policy\(^{28}\). The reasons mentioned in the study by Mayers and Smith (1982) explain why the firm foregoes self-insurance and the reasons analysed in Mayers and Smith (1981) hint\(^{29}\) toward why the firm does not pay the managers and let them organise the insurance by themselves.

The comparative advantage of the specialised insurance firms in processing claims, evaluating costs and managing legal disputes may be a good reason to buy an insurance policy, see Mayers and Smith (1982) and Skogh (1989). Moreover, as Mayers and Smith point out, the conflict between owners and managers may explain insurance policies as a monitoring device. If a specialised insurance company has a comparative advantage in monitoring risks and counter measures, the owner may delegate the supervision and control of managers to the insurance company\(^{30}\).

In addition to fire or damage insurance, business-interruption insurance policies may also be used to control incentives and thus used as contractual devices. Mayers and Smith (1982) write:

\[
\text{... occurrence of an unusual circumstance can alter incentives. Suppose a fire destroys a large part of a single plant of a multi-plant corporation. The firm might have an incentive to reduce employment, violating the anticipated allocation of risk bearing. Business-interruption insurance covering ordinary payroll would control the incentive. In this case, by guaranteeing the contract the reduction in the workers’ reservation prices can be sufficient to cover the loading fee for the insurance policy.}
\]

The text continues (footnote on page 288):

\(^{27}\)Parry and Parry argue wrongly that in the previous analysis the firm has been risk-averse. In the papers by Mayers and Smith (1981 and 1982), the firm consists of different interest groups. As a consequence, preference of the firm is not a meaningful concept.

\(^{28}\)The authors use director and officer (D&O) liability insurance as an example. It is designed to cover the errors and the omissions of board members and company officers.

\(^{29}\)Grouping of policies is more likely if the information asymmetry between purchaser and seller of insurance policy is high, when employers’ incentive to monitor the employees’ purchase is greater and when the demand for insurance is homogeneous in the group.

\(^{30}\)For example, an insurance company may control the condition of a sprinkler system. By buying an insurance policy, the management may credibly promise to maintain the system.
This appears similar to the risk-shifting incentive discussed earlier, but the rationale is different. In this case the firm can provide the optimum level of risk for the contract.

The authors do not elaborate the idea further. For review of other reasons to buy a business-interruption insurance policy, see Järvikare (1996).

Mayers and Smith also have another idea on how insurance can be used to control incentives. They have presented (Mayers and Smith, 1986) an analysis of a firm that is partially financed by debt. If the casualty loss is large enough, the value of the firm, even after reparations, may sink below the value of its debts. In such a situation, the shareholders wouldn’t be ready to invest in the rebuilding of the damaged asset. However, by actuarially fair insurance the incentive problem can be solved and the rebuilding guaranteed. In this context, the insurance contract protects the debtor.

2.3.3 Trade union models

During last two decades, trade union models have become common in analysing labour markets, especially in Europe. In these models, wages are set by a trade union or wages are negotiated between a trade union and a firm or between a trade union and an employers’ association. One may ask if trade union models can be used to explain observed insurance demand by the firms.

In the literature of trade unions, three main types of models are analysed: monopoly union models, efficient-bargaining models and the so-called "right-to-manage" models\(^{31}\), see surveys by Oswald (1986) and by Ulph and Ulph (1990). The most important endogenous variable in trade union models is the wage level. It is an important factor determining profits. If wages can be down-sized in less uncertain demand conditions, the firm would want to buy an option, a kind of an insurance policy, to sell at constant prices just to lower the wage level. If one, however, concludes that a decrease in uncertainty does not cause a decrease in the wage level and a corresponding increase in the profit level, there is no reason to purchase an insurance policy.

\(^{31}\)In a monopoly union model, the union decides the wage, but takes into account the labour demand schedule of the firm when maximising utility of the workers. In efficient contract models, both wage level and employment are agreed upon by the labour market parties. Because this kind of agreements do not exist in the real world, the "right-to-manage" models, where the wage is agreed upon by the labour market parties and the level of employment is decided unilaterally by the firm, have become more popular recently.
Consequences of uncertainty are, indeed, studied in the context of the trade union models. The classical reference is Oswald (1982), in which a monopoly union model is studied. The similarity between Oswald’s model and the model presented here allows us to compare quite easily the results obtained\(^\text{32}\). Oswald shows that the effect on the union’s optimal wage of an increase in the riskiness of product demand, i.e. price level, is ambiguous. With one particular form of production function\(^\text{33}\), he is able to show that there is no effect.

Oswald studies also other sources of uncertainty. He shows that an increase in the riskiness of unemployment benefit decreases the union’s optimal wage. Also, if the employed workers show non-increasing absolute risk-aversion, a rise in the riskiness of the income subsidy increases the union’s optimal wage. The effect of the unions’ optimal wage of an increase in the riskiness of the comparison wage rate is ambiguous.

In can be concluded that in the Oswald’s trade union model the firm has no clear motive to reduce uncertainty and thus the model does not explain why firms buy insurance policies.

Burda (1995) shows that many aspects of observed trade union behaviour can be understood in the light of the monopoly union model. In his analysis, the monopoly union provides insurance against wage variability via its wage policy. An optimal contract is characterised by real wage flexibility only in response to systematic changes in uncertainty. However, like the analysis of Oswald, the model is not capable of explaining insurance purchases as the firm is able to insure the workers’ incomes by itself.

### 2.3.4 Implicit contract theory and beyond

According to the implicit contract theory, a firm provides more certain income for the workers than pure production conditions would imply. This is a welfare improving arrangement if the firm tolerates risks better than the employees, see the classical papers by Baily (1974) or Azariadis (1975). Because explicit contracts of risk sharing between the firm and the workers are rare, the relation is assumed to be implicit. In theoretical models, the contracts specify the amount of the labour to be utilised and the wages to be paid in each state of nature.

\(^{32}\)In the model of Section 2.2, uncertainty was modelled as a change in production capacity. Technically, this corresponds to price uncertainty in Oswald’s basic trade union model. In both models, an uncertain element enters into the production function as a multiplicative factor.

\(^{33}\)He assumes that \( F(N) = (1/(1 - \gamma)) N^{1-\gamma} \).
One may wonder whether the existence of firm insurance policies can be explained by implicit contract theory. However, it should be noted that if the firm is risk-neutral, it would be able to bear all the risks and there would be no reason to buy a policy\textsuperscript{34}. It is, nevertheless, interesting to compare the basic results of implicit contract theory with the results of the analysis of Section 2.2.

The model in Section 2.2 was similar to that of the implicit contract studies, but was solved by assuming a constant wage. The implicit contract theory outcome of the optimality of constant wage supports indirectly the assumption made in this study. But, as Manning (1990) points out, the wage rigidity result depends crucially on the assumption that the employer is risk-neutral.

The main aim of the implicit contract theory was to explain the existence of unemployment. The theory is good in explaining why wages are rigid, but it is not, however, very successful in explaining unemployment as such. The basic models predict higher employment than would prevail in an auction type of labour market because the firms hoard labour during a recession\textsuperscript{35}. It is interesting to note that in the model of Section 2.2, there is a similar property. Reparations are made even though they are non-profitable and consequently production is inefficiently high.

The over-employment result has been influential in creating pressure to find a better way to understand implicit risk-sharing contracts between the firm and its workers. As in many other areas of economics, the asymmetric information models have been useful in solving remaining theoretical puzzles of implicit contracts. If the information sets of the firm and the workers are different, i.e. if the parties do not observe the same facts, it is possible to construct models where underemployment exists. A paper by Calvo and Phelps (1977) was one of the first analyses where it was assumed that the workers cannot observe the values of uncertain variables\textsuperscript{36}.

The papers with asymmetric information were able to explain the existence of unemployment. However, the papers lost an important property of the original papers by Azariadis and Baily: in models of asymmetric infor-

\textsuperscript{34} Things are seldom that simple and there are many reasons why the firm cannot provide itself the insurance that it would like to offer its employees. Firms may be risk-averse or all kinds of rigidities and contractual difficulties may hinder the operation of market forces. Capital stock of the firm also limits its capability to carry down-side risks. These explanations are already mentioned and have nothing in particular to do with the basic ideas of implicit contract theory.

\textsuperscript{35} The over-employment results was first pointed out by Akerlof and Miyazaki (1980).

\textsuperscript{36} The other early papers are Grossman and Hart (1981) and Green and Kahn (1983).
mation, an unemployed individual who would like to swap places with an employed person with the same characteristics does not exist (Oswald 1986). As Oswald points out, the models based on the assumption of asymmetric information assumed that workers cannot rely on the government’s unemployment benefits when unemployed. In the classical papers by Azariadis and Baily, this assumption was made. Oswald returned the assumption for empirical grounds into the asymmetric information implicit contract models. He produced a model where both desired properties were present: there is under-employment and involuntary employment exists, i.e. the unemployed workers envy the employed workers (Oswald 1986).

The problem of enforcing the contract was dealt with in Section 2.2. The strong individual rationality constraint implies that not all contracts are implementable. The same topic has been discussed also in the context of contract theory. Holmström (1981) argued that reputation effects or long-lasting contracts may solve the enforcing problem. Workers or firms that are known to be contract-breakers would earn a bad reputation and this threat would curb down the endeavour for the short-run gains. Also, the contract may include seniority wages. A promise of a higher wage in the future may be good enough reason to stay in the job even if the market wage is higher than the wage specified by the contract. Based on this, it can be argued that an insurance policy purchase may not be very important for old institutions with a long history and that the same contractual effect can be obtained by reputation formation. One may wonder whether this is an additional reason why public sector institutions and large corporations do not insure. The possibility to pool risks in-house is, of course, the main reason why insurance policies are not bought by large institutions.
2.4 Discussion

In this chapter of my thesis, it was shown that the risk-neutral firm may obtain a higher expected profit level by buying an insurance policy. This may be the case if the firm can deviate from the agreement if a fire or some other accident makes production non-profitable. Moreover, it is required that the reparation costs exceed the value of expected production after preparations.

By an insurance policy the contracting parties are able to create suitable private incentives that serve social purposes. Though production with low capital may be non-profitable from the firm’s point of view, it may well be profitable from the workers’ point of view. Continuation of production requires, however, reparations which are, by assumption, non-profitable for the firm.

In the real world, an insurance policy may be useful if without it the firm cannot credibly promise to build the factory again or to repair the machinery after a major damage or a fire. This may be the case if the factory is old or if, for some other reasons, the reparation costs are high compared to value of production. If production is socially profitable, then there is in principle a possibility to share the rent obtained by repairing the factory and continuing production. In general, the existence of mutually profitable production possibilities does not guarantee that production can ever be continued, but in the model analysed in this chapter, an insurance policy would solve the incentive problems and the socially optimal production level can be maintained.

The analysis was based on a simple model of implicit contracts where the agents were risk-neutral. It was shown that if the firm is able to cease production in adverse states of nature, the efficient contract cannot be maintained, but an insurance policy providing the minimum capacity for (privately) profitable production is welfare improving. However, it is possible that the minimum capacity for profitable production cannot be provided when the realisation of the productive capacity is too low, i.e. if the damages are too large.

The assumption of constant wage should be discussed briefly. The model used in this chapter was constructed by assuming constant wages, i.e. it was assumed that the wage $C_1$ and the unemployment benefit $C_2$ are not conditional on the realisation of $\theta$, productivity.
One may ask why the contracting parties do not decide that wages are adjusted according to production conditions to provide a positive profit level in all circumstances. In principle, risk-neutrality of the workers allows them to provide insurance for the firm. This possibility can be neglected on the following grounds: first, the workers’ ability to provide insurance for the firm is a theoretical artifact produced by the extreme assumption of risk-neutrality. The risk-neutrality assumption was made to exclude from the analysis all "conventional" motives to insure, but not to describe the real world. Secondly, profits vary much more than wages at the micro level. Hence, even if the workers provide insurance, insurance coverage must be very limited. Third, insurance companies must have higher productivity in processing information than the workers. Thus the informational asymmetries outside the analysis justify the assumption that the insurance is provided by the specialised companies.

I have argued that an insurance purchase can be used to reduce inefficiency caused by contractual inefficiencies. The question of possibility of using conditional wages to overcome the individual rationality constraint raises an alternative interpretation for the firm insurance policies. Insurance policies can be useful in reducing the utility loss caused by unconditional, i.e. non-flexible, wages.

Bankruptcies in the event of a severe fire or an accident are not the only cases where non-flexible wages cause utility losses. In many cases, disturbances are of macro nature and it is thus quite difficult to imagine an insurance policy protecting the continuity of production. However, the use of price options as well as foreign exchange futures and forward contract have become more popular recently. The ideas presented in this chapter are, perhaps, useful in understanding why firms that should be risk-neutral, are actually proceeding in that direction.

The nature of the analysis should be kept in the mind when planning empirical testing of the model. The quite restrictive assumptions that are made are justified, because the analysis aimed to show that there are other motives than risk-aversion that could explain why firms buy insurance policies. I do not intend to claim that controlling incentives is the only reason of purchasing an insurance policy. It is, perhaps, one motive among others. A proper empirical test of this theory should take this reservation as a starting point.

At first glance, the properties of real insurance contracts coincide quite
well the obtained theoretical ones. The model predicted that there is no guarantee for the maximum production capacity, this seems to fit quite well with casual observations. It is also interesting to note that in the model, very large damages are not covered by insurance policies. This seems to correspond to insurance policies in the real world.

Without any empirical study\footnote{At the moment, there are no empirical studies available for fire insurance demand. It is, however, interesting to look empirical results by Järvikare (1996) on business-interruption insurance demand. Using data obtained from a leading business interruption insurance company, he gets the following results. Insurance purchases are increasing in the size of the firm, but the elasticity is less than one. Small, widely-owned firms buy more insurance than other firms, and large ones less.

Järvikare’s results indicate that other motives than the pooling of risks explain insurance purchases. However, as firms get bigger, closely-held and family-owned firms purchase insurance policies with larger coverage. The interpretation is clear: when risks become bigger, the traditional motives to buy insurance policies override the contractual motives.} it would, however, be premature to conclude that the model corresponds to reality. A proper test of the theory should start from the very nature of the theoretical model. In the model, insurance policies were used as a tool to grant credible promises to the workers. The same mechanism would apply to the sub-contracts. Hence, one may expect that labour-intensive firms or firms with many sub-contractors are active to insure themselves.

Moreover, because reputation effects may serve the same purposes as insurance policies serve, one would expect that new and small firms buy insurance policies eagerly. Old and large firms are in a better position to develop a good reputation and do not need a third party to guarantee promises. Unfortunately the possible observation that old and large firms do not buy insurance would have an alternative interpretation. Large firms are able to pool risks inside the company and old firms typically have stronger balance sheets than new firms. Both factors explain why insurance policies are not needed.
Chapter 3

Childcare and wage distribution

3.1 Introduction

The aim of this chapter is to offer a theoretical explanation for gender wage differentials and labour market stratification. The starting point of the analysis is the fact that women are more likely than men to be responsible for childcare. This means that it is costlier for a female employee than for a male employee to commit oneself to work in a way satisfying the employer. The widely discussed labour market flexibility on one hand and rigid demands of children’s well-being on the other are difficult to combine. However, this difficulty is not constant over time or across societies. It depends, \textit{inter alia}, on the availability, quality and costs of childcare facilities.

The basic idea of the interplay between domestic and market activities is not new, see Becker (1993). Becker emphasises the importance of specialisation in the allocation of time and in the accumulation of human capital. Married women specialise in childcare and married men in market activities\textsuperscript{1}.

In the study at hand, we go deeper into the problem. It is argued that the most important element is not the required labour input as such, but the flexibility and the commitment that it requires. This kind of service is very difficult to buy from the market and there is, hence, a trade-off between the commitments of the parents. The requirements of flexibility

\textsuperscript{1}The problem with Becker’s analysis is that it does not explain why childcare is still performed in the family. Organising it in specialised units would allow much more effective specialisation.
and commitment order jobs, firms and professions in the labour market. Professions with high commitment are well paid because the required effort is very difficult to substitute with hours worked. Using an explicit model, it is argued that individuals with high domestic commitments are allocated to low commitment, low pay professions. The model is then used in discussing the possibility to change family roles.\footnote{The study by Francois (1998) has a very similar starting point. However, gender roles are endogenous in his study, see Subsection 3.1.7.}

### 3.1.1 What is special in childcare?

What is special in the care and supervision of children? What makes it different from other domestic duties? There are two important elements.

First, in all human activities there are elements which are not marketable. In childcare, this element is large and it is essential in the child-parent relationship. It is very difficult to hire somebody to tell the same stories to your child as you like to tell them. Only parents\footnote{The definition of family differs across cultures. Here parents should be interpreted very broadly; it may also mean other members of the family.} can take full responsibility for the health of their children. What is important from the gender point of view is that this form of labour input is strongly complementary to the other element of caretaking. For example, feeding and supervising go hand in hand. Similarly, the daily hygiene and looking after the health of the child are easy to combine. Hence, it is common for the person who is responsible for daily routines with children also to be responsible for non-market activities.

The second point is that the time span of children is short and thus supervising and care activities are needed on a daily basis. The laundry can wait, but children cannot.

Taking care of one’s child requires a lot of the same kind of labour input as any effort at work. As a parent of a child, one is supposed to be ready to wake up in the middle of the night if needed. As an airline pilot one must be willing to stay up even when the plane is lagging behind the timetable. Imagination and commitment are as valuable at home as in business or in politics.

The very fact that women have a heavy domestic work load affects their situation in the labour market\footnote{See Iïmakunias (1997) for a discussion on conventional labour supply consequences of different childcare arrangements.}. However, the basic argument of this chapter is more elaborated. It is argued that women are generally more
3.1. Introduction

responsible for childcare than men and that the basic nature of childcare is such that it is quite difficult to combine it with professions characterised by high effort requirements. Professions that do not require total concentration or commitment are more easily combinable with heavy childcare loads.

3.1.2 Domestic effort versus effort at work

Worries of everyday life are easier to bear if the employer does not demand full concentration at work. Shorter working hours allow more time for cleaning, cooking or car reparations. Hence, there is a trade-off between professional and domestic tasks.

Efficiency wage models are a widely known application of a similar idea. The distinction of hours versus effort is not, however, the only possibility of aggregating heterogeneous labour inputs. In fact, these terms may be interpreted in various ways. Sometimes effort refers to intensity or pace of work, sometimes it is only an index of the quality of workers’ performance. In this chapter a very specific definition of effort is adapted. The term effort refers to a worker’s commitment to serve the employer’s goals flexibly. This kind of effort is important in continuous production processes and production processes with indivisibilities (Ramaswamy and Rowthorn, 1991). As a consequence, the term labour refers to everything that may be agreed on beforehand. The reason for using such a definition comes from above analysis of domestic duties.

It is reasonable to assume that workers demand better hourly compensation if they are asked to work harder. Because domestic duties vary, compensation demands of workers in different occupations also vary. For example, unexpected overtime work is much more difficult to cope with for a single mother than for a bachelor. Hence, it may be expected that a person with strong family commitments demands a higher compensation for a job than a bachelor if the job requires irregular working hours, travelling and a full commitment to the targets of the employer.

As a consequence, we may expect that individuals with strong family commitments are allocated to jobs where the effort is less important. Also, if pay is positively related to effort, then these people are also less paid. At least at first glance, these features are characteristic for the situation of women in the all modern societies. However, a more careful discussion of empirical implications is postponed until the end of the chapter.

It should be noted that actual labour market outcomes may not fully reflect demand and supply conditions of effort. A market for effort may, as well as the market for conventional labour, be noncompetitive. It is,
however, unlikely for market imperfections to totally distort the underlying
tendencies. Hence, the analysis is carried out assuming perfect information
and competitive markets. The only exception is that the number of firms
in the formal analysis is assumed to be constant.

3.1.3 Determinants of female childcare load

The analysis in this chapter is based on the notion that in modern societies
the time that childcare requires is not the issue. This can be justified by an
example. Assume that there are \( k \) women seeking daycare for their children.
They always have an option: those \( k \) women may form a collective, where
\( k-1 \) women work at the labour market and one takes care of all the children.
Although the double taxation causes extra costs in the form of a tax wedge,
the arrangement should be profitable if \( k \) is large enough. However, this
kind of shared childcare is not common in industrial countries.

The problem is that the required care is very difficult to produce outside
the family.\(^5\) Market labour or an arrangement of the above-described na-
ture are not very good substitutes for parental care. Though the required
working hours are marketable, the required effort is not necessarily. There
is a risk of private childcare arrangements outside the family not reduc-
ing, or redistributing the cost of effort performed at the labour market well
enough. Either children or performance at work suffer.

Technological development has increased productivity also in childcare
production. In addition to better nutrition, housing, health care and other
factors affecting the physical safety of children, various immaterial elements,
education and communication technologies included, have made it possible
to increase the standard of living of children at the same time when the
childcare load has decreased. In general, the benefits of childcare produc-
tivity growth have been distributed between children and parents. It seems
that women have benefited the most in this process. Decreased number of
births in an average woman’s life is an important factor in this development,
but also effort costs per child have declined substantially over the decades.
What is most important from the policy perspective: decreasing the female
childcare load is quite a generally accepted policy target in industrialised
western countries. Moreover, it is possible that there are purely economic
reasons for providing childcare. For example, firms having female workers
may benefit from the availability of public childcare. However, as Rosen

\(^5\)François (1998) expressed this idea by assuming in his model that there is family
specific human capital such that the members of the family are more productive in the
household production than outsiders.
(1995) argues, childcare programmes are expensive\textsuperscript{6} and require heavy taxation and are thus also causing substantial welfare losses.

I would like to argue that there are three alternative approaches to reducing the female childcare load further. The load can be divided more equally in the family, reduced by further decreasing the number of children, or part of the load can be borne by the government. Governments can do this, for example, by providing public daycare facilities. The latter option is discussed first.

### 3.1.4 Public daycare

Public daycare facilitates participation in the labour force. As the example given in Subsection 3.1.3 suggests, private arrangements can do the same. However, the public alternative is superior in facilitating a strong supply of effort as well, as the main problem in combining family and working life is not the availability of childcare, but merely quality of it. Moreover, the price of the service matters. For example, though the actual services are sometimes produced by charities or other non-profit organisations, daycare is almost totally financed by the governments in Nordic countries\textsuperscript{7}. Because of subsidies, the staff-to-children ratio is higher in public daycare and the principal-agency problems are less severe than in private services as care production is supervised and controlled by the authorities.

Providing education is, perhaps, the most important government intervention in family life. In addition to formal education, schools are important in transmitting cultural values and in providing care and protection for the children. Basic education is organised by public authorities in almost all countries, but very often no special public daycare is available. In many countries, if daycare facilities exist, they are private and quite expensive. However, it is common for children to start school at quite an early age and thus primary schools provide a kind of public childcare service in those countries. Specialised childcare is, however, superior in reducing effort costs of childcare for the following three reasons.

\textsuperscript{6}The total cost of public daycare programmes in Finland was FIM 6.8 billion in 1992. This is 1.4 per cent in relation to GDP.

\textsuperscript{7}In 1987, parents’ share of the daycare running cost was 18.3 per cent in Denmark, 16.3 in Norway and 10.8 per cent in Sweden, see Leira (1992). In Finland, parents’ share was 14 per cent in 1992 (STM 1994). Because of the high subsidies in the public childcare system, a private alternative has almost disappeared in the Nordic countries. For example, the share of Swedish children in the age group of 3-6 years attending government-sponsored daycare was 81 per cent in 1989 (Leira 1992). The corresponding figure for Finland was 53 per cent in 1992 (STM 1994).
First, the way childcare is organised is better from the working life point of view. For example, timetables of the Finnish daycare centres are designed to suit working hours and there is normally only a short summer break. In this respect, a combination of short working days and long holidays, typical for the primary schools in all countries, make them inferior in providing care.

Second, though formal schooling starts at an early age in many countries, daycare starts even earlier. For example, Finnish children may go to a nursery at the age of one year\(^8\). As a consequence, the break in women's working career may be much shorter in countries providing daycare than in others. Duncan (1996) compared the availability of public childcare services in European countries in 1989. According to him, the ratio of childcare places to the number of children does not vary very much between countries in the age group from 4 years to the age of starting school. In the age group of 0-3 years, however, variation was substantial. In East Germany, Sweden, France and Belgium, the ratio varied between 20 to 80 per cent, but in the other European countries covered by the study\(^9\), the share was negligible.

The third difference is in staff and the premises of establishments. Schools are planned to provide for education, but daycare centres are more capable of providing for care. Children have several meals, they may take afternoon naps and playing facilities are better in the daycare centres. Hence, no private arrangements are needed even when both parents of young children want to work full time. Hence, it is justified to conclude that daycare and schooling are substitutes, but only imperfect ones.

The basic "ideology" of daycare is totally different in Nordic countries from many other European countries. Although educational goals are important also in the Nordic model, these dominate in the other countries. In the countries without public daycare services, almost all women are outside the labour force when their children are young. Moreover, during the school years of their children, women are more often employed in part-time jobs than, for example, in the Nordic countries. The services available do not allow wives to commit to work effectively outside the family unless the husbands make radical changes in their lifestyle. It is clear that the trade-off between domestic duties and labour market commitments is much more visible and tight in those countries than in the countries where public

---

\(^8\)This age is now 12 months, but was previously 8 months. Local governments are obligated to offer a place in a nursery or a publicly supervised family care for all children under three years of age.

\(^9\)The group consists of West Germany, the Netherlands, UK, Italy, Luxembourg, Ireland, Greece, Portugal and Spain.
childcare is provided.

It should be emphasised that the picture of the Nordic countries drawn here is necessarily simplifying. For example, as Leira (1992) argues, childcare policy in Norway deviates from that of the other Nordic countries. Moreover, Ilmakunnas (1997) shows that there are labour supply reducing elements also in the Finnish family policy. She analysed the effect of Finnish child home care allowance\textsuperscript{10} for women's labour supply decision. According her results, introduction of the allowance increased popularity of private care only marginally, but the effect on popularity of maternal care was substantial\textsuperscript{11}. By having such a effect, the allowance is a major exception in the Finnish family policy, which has traditionally targeted increased female labour supply.

3.1.5 Changing family patterns

The female childcare load can be reduced without any public intervention by either decreasing the number of children or by dividing the load more equally within the family. The former is the only solution if female labour force participation rises without major changes in gender roles and if there are no public childcare facilities available. In that case, children are often concentrated into few families\textsuperscript{12} with many children. Modern Italy is a good example of this kind of development. Availability of childcare is very limited and gender roles have changed only marginally. As a consequence, birth rates are currently very low in Northern Italy and as the rates are declining also in the South, the average figure is one of the lowest in Europe. It is unclear how sustainable the current combination of high female participation rate at the labour market and low birth-rate can be.

Hence, it is not very surprising that birth rates are high in countries with generous public daycare\textsuperscript{13}. Moreover, Ilmakunnas (1994) has shown

\textsuperscript{10}This is quite a generous income transfer paid to all families with children under the age of three who do not use public daycare services, but rely instead on parental or private care. It was introduced in the late 1980's.

\textsuperscript{11}Using micro-data from Helsinki, she estimated that rising allowance from zero to FIM 3 500 (about the level of actual benefit) would increase mother's probability of staying home from 35 to 55 per cent.

\textsuperscript{12}Often these are migrant families from less developed countries.

\textsuperscript{13}According to Eurostat (1995), between the mid-1960's and 1992, fertility in the EU was in constant decline. The overall pattern was affected by trends which varied from one Member State to another. There was a slight recovery in the fertility rate in Belgium and the Netherlands and a greater recovery in Denmark, Luxembourg, Finland and Sweden, while the rate was unchanged in Austria, France, Germany and the UK. The southern
that there is a positive cross-country correlation between birth rates and GDP-share of children-related public expenditures and transfers.

The latter alternative, i.e. changing gender roles, might be more attractive if a public policy action has already taken place and has reduced the costs associated with childcare (Francois 1998). However, in the model analysed in this chapter, it will turn out that public policies would not have any effect on decision making when gender roles are allocated in the families. In what follows, the analysis of Francois is reviewed in Subsection 3.1.7 and commented in Section 3.4.

3.1.6 Traditional gender studies

It is worthwhile to try to place this study in the existing literature. Rubery (1987) divides neoclassical explanations of women's wages into three broad categories. The first one is Becker's discrimination hypothesis. According to this hypothesis, women are paid less than their marginal product because employers or fellow workers do not like them as employees or co-workers. In this sense, female workers create a negative externality.

The second hypothesis is based on human capital theory, see Becker (1993). For various reasons, mainly because of self selection, women tend to have less work experience and a lower education level and hence their productivity is lower than the productivity of men. Women's role in household and childcare is a barrier to the accumulation of human capital. Hence, low human capital is a mediating factor, not the fundamental cause of the lower pay of women.

Rubery's third category, job segregation, consists of three different families of theories. The first of these segregation theories notes women's role in reproduction. According to this explanation, women are allocated to less paid jobs because they are less committed and less productive workers. In this sense, this theory is quite similar to the human capital theory mentioned above. However, the mediating factor is not human capital but allocation of well-paid jobs. The second segregation theory is based on prejudice about which jobs are suitable for women. The third explanation of labour market segregation is based on postulated real differences of genders in production.

Rubery (1987) also notes some non-neoclassical explanations of different pay. Although this field of research is wide and dispersed there is still something in common. All these theories are based on the assumption that

---

EU countries (Greece, Spain, Italy and Spain) experienced a sharp, swift drop.
3.1. Introduction

The fundamental cause of women’s low pay is their inferior position in the social and cultural system. Patriarchal theories of women’s wages are based on the fact that women’s position was inferior prior to capitalism and capitalism had to adapt to this structure. Women should earn less than men, otherwise the patriarchal structure of society breaks down. The market segmentation approach explains lower pay of women by referring to the usefulness of dualism of the economy. The extreme version of this theory states that the purpose of segmentation and divisions of the labour force is to minimise the power of workers. The family-wage approach takes into account the fact that labour income is divided inside the family. According to this approach, it is not correct to analyse an individual worker in the labour market because his or her situation is heavily affected by other members of the family.

This chapter definitely belongs to the job segregation theories. Genders are allocated to different jobs because they face different costs of working efficiently. However, the idea behind the model has some links to the second explanation category mentioned by Rubery. Women work less hard and are less committed to work than men. This is not due to the lower technical productivity of women, but due to market allocation. It is interesting to note that there is no discrimination in the model and the outcome is "fair" because employers get exactly what they pay for.

3.1.7 Exogenous or endogenous gender differences?

The recent paper by Francois (1998) is an interesting challenge for the traditional gender studies as well as for the analysis presented in this chapter. In his model, all gender differences arise endogenously. The author does not deny the possibility of inherent gender differences but argues that discrimination itself does not need to depend on such differences. He demonstrates that discrimination may persist even if inherent differences between men and women become less pronounced and indeed disappear altogether.

In Francois’ model, labour markets are competitive, firms profit maximising, women and men are identical and neither employers nor anyone else has a preference for men or an exogenous propensity to discriminate. The analysis is based on two gender-symmetric assumptions. First, women and men are organised into two-person households, in which both are equally well-endowed with household-specific human capital; they are able to provide households services for themselves and each other at a cost lower than
the external market rate. Second, the labour market consists of two job
types. There is permanent over-demand for efficiency wage jobs in which
workers receive a wage premium in order to deter shirking\textsuperscript{14}. There also
exist non-efficiency wage jobs which pay no such premium.

The features of Francois’ model imply that if and only if one member
of a couple alone has an efficiency wage job, both can benefit by a trade
which increases the household work of the low-wage member in return for
money. The trade decreases household work of the high-wage member of
the family and the employer can lower efficiency wage payments while still
satisfying the worker’s participation constraint as threat of job loss entails
loss of gains from trade in addition to loss of the wage premium. Herein lies
a firm’s desire to discriminate. If all other firms discriminate and allocate
attractive efficiency wage jobs to one sex only, then a single firm can ensure
its employees receive this benefit, and thus can pay a lower efficiency wage
only if this firm employees men exclusively as well. Only by doing so can
it ensure that its employees come from households where they are the sole
efficiency wage recipient, and therefore they can enjoy the benefits of a trade
within household. Hence, discrimination is a Nash equilibrium.

Francois argues that policies focusing explicitly and solely on the labour
market may alleviate a substantial amount of discrimination, even though
external factors remains unchanged. For example, the author is able to
show that a policy of affirmative action, even when targeting only a part
of employers, can bring an end to discrimination. Furthermore, policies
that target gender differences directly, such as childcare subsidies which
reduce female costs of employment, are markedly less effective in shifting
the economy from a discriminatory equilibrium. The policy advises drawn
from Francois’ analysis are compared with our results in Section 3.4.

3.1.8 Content and organisation of the chapter

The basic model with two labour inputs (effort and conventional labour) is
introduced in Section 3.2. In the same section, the operation of the model is
illustrated assuming two types of firms, but only one gender type. Section
3.3 extends the analysis and studies the allocation of genders over firms.

Despite substantial inertia, family roles are not biologically or psycho-
logically fixed. Individual families are able to alter their internal division of
labour and effort. Determinants of this kind of family role modernisation
are discussed in Section 3.4. Section 3.5 concludes the analysis.

\textsuperscript{14}This feature is adopted from Bulow and Summers (1986).
3.2 Basic model

Wage formation and labour market stratification are analysed using a simple model with two labour input components. The labour demand side is similar to that in Ramaswamy’s and Rowthorn’s studies\textsuperscript{15} (1991,1993).

3.2.1 General structure of the model

The model of this chapter has a similar structure as the model in Chapter 2. The basic elements are labour demand derived from the firm’s maximisation problem and preference relation of the workers. The preference relation is again determined according to individual preferences. Despite the similarities, the model is solved differently; the firm is assumed to be capable of setting the wage level\textsuperscript{16}. However, the labour market is structured in such a way that there may be excess demand for high productivity labour, i.e. the labour input of ”high quality” workers. As a consequence, not only the firm, but also the ”high quality” workers extract some rent.

The firms produce goods using labour and effort and the workers are paid according to their effort. The firm decides on the wage level taking into account how the wage level affects performance of the workers. Mathematically, the problem of the firm is

\[
\max_{N, W} \Pi(N, E(W), W)
\]

\[
\text{s.t. } U(E, W) \geq U,
\]

where

\[
\Pi(N, E(W), W) = F(N, E(W)) - WN
\]

is the profit function. The firm sets the number of workers \( N \) and the wage level \( W \) to achieve the highest profit \( \Pi \) available. The price level is assumed

\textsuperscript{15}They generalise the well-known Solow condition and show that a firm for which the effort level is relatively more important pays higher wages than a firm for which effort is of less importance. See also Stiglitz (1987).

\textsuperscript{16}This extreme assumption is justified because the aim of the analysis is to link production and domestic activities together. The negotiation power of different types of workers and effect of the trade unions are abstracted away to make the basic argument as clear as possible.
to be fixed. If the link between wage and effort is bypassed, then profit is increasing in effort $E$ and decreasing in wage $W$.

The opposite is true for the workers; their utility is increasing in wage and decreasing in effort. The utility realised in production, i.e. $U(E, W)$, must exceed the alternative $\overline{U}$, i.e. utility prevailing outside of the firm. Otherwise the workers will not participate in production.

The restriction $U(E, W) \geq \overline{U}$ can be interpreted alternatively. The participation constraint implies that efficient and flexible working practices should be compensated by higher wages if utility is fixed. Assume that $U(\cdot, \cdot)$ is well-behaving, that it can be inverted and treat utility as fixed, i.e. set it to $\overline{U}$. Then one obtains from equation $\overline{U} = U(E, W)$ an alternative description of the preferences of the workers.

$$E = E(W, \overline{U}).$$

This is the effort function. For any given utility level, effort and wage are positively related. Similarly, for any fixed wage, the higher the effort, the lower utility obtained. The effort function defines the shape of iso-utility curves for the workers. All points $(W, E)$ satisfying Equation (3.2) are on the same utility contour.

Effort functions of the form (3.2) are determined by the worker’s outside option. In the most simple case of our analysis, the outside option is the utility of an unemployed worker. In more complex situations, utilities of workers employed by other firms should be taken into account.

In the traditional efficiency wage model the production function is

$$Y = F[E(W)N],$$

where $E(W)$ is efficiency as a function of wage and $N$ is quantity of labour input; $E(W)$ and $N$ enter into the function multiplicatively. The firm maximises profit $\Pi = F[E(W)N] - WN$ and sets the wage $W^*$ such that

$$\frac{W^*E'(W^*)}{E(W^*)} = 1.$$  

This is the Solow condition. The effort function that links together the efficiency of labour and the wage level is increasing.\footnote{Efficiency wage models, i.e. models with a positive relationship between effort and wage, have their roots in development economics, where there is a strong belief in the}
3.2. BASIC MODEL

In this study, effort \( E \) is a measure of the performance of the workers. Interpretation of effort is closely related to flexibility. The same amount may be produced by using many hours in an inflexible way or few hours very efficiently, just-on-time. The number of workers represents the quantity of labour used in production. In this respect, the model is in line with the traditional efficiency-wage models where effort and hours are combined. However, the production side of the model is richer in detail than production in the basic efficiency-wage model because it is necessary to allow the importance of effort to vary across the firms. As a result, the Solow condition does not hold anymore and the tools developed by Ramaswamy and Rowthorn (1991, 1993) are very useful.

Though the structure of the model analysed in this chapter is described by (3.1), the analysis is actually carried out with the firm’s iso-profit curves and the workers’ effort functions. The model is solved noting that in the optimum, firms select the highest iso-profit curves that are consistent with effort functions.

To facilitate the analysis, the production function of the model is specified in such a way that the iso-profit curves are linear in logarithmic wage and logarithmic effort space\(^{18}\). Hence, all combinations of \( e \equiv \log(E) \) and \( w \equiv \log(W) \) satisfying

\[
e = \text{constant} + \lambda_i w
\]

result in the same profit. \( \lambda_i \) is a firm-specific parameter describing the importance of effort in production. Preferences of the workers can be explained relation between real wage, nutrition and productivity. In modern industrial countries, this explanation is quite irrelevant. More recent efficiency-wage models range from imperfect information theories to so-called sociological explanations. The cornerstone of the Shapiro-Stiglitz model (Shapiro and Stiglitz 1984, see also previous articles by Calvo 1979 and Salop 1979) is the fundamental disability of the firm to observe the workers' effort. In the absence of proper screening devices, the firm pays high wages to the workers and fire them immediately if they are caught shirking in random checks. The threat of losing a well-paid job deters workers from shirking and increases the average effort. In other models with imperfect information, the relation between wage and effort is based on analysis of search behaviour. Offering a higher wage will reduce the quitting rate (Salop 1979) or will lead to a better pool of applicants. In the so-called sociological models, the effort wages result from the notion of fairness (Akerlof 1984). Reduction in wages by a firm may be considered unfair by the workers, leading them to supply less effort.

\(^{18}\)The analysis of the production side of the model is the topic of discussion in Subsection 3.2.2.
pressed in the same space. It is shown in Subsection 3.2.3 that reasonable assumptions produce a concave relationship between effort, wage and utility.

\[ e = g(w - U), \quad g'(\cdot) > 0, \quad g''(\cdot) < 0. \]  \hspace{1cm} (3.6)

In the optimum \( \lambda_i = g'(\cdot) \) and thus the optimal effort and wage in the firm \( i \) depends on the parameter \( \lambda_i \), i.e. on how important effort is in that firm\(^{19}\).

### 3.2.2 Demand for labour and effort

The production function allows the importance of effort to vary across the firms.

\[ Y = F(E^{1/N}), \]  \hspace{1cm} (3.7)

\[ F(E^{1/N}) > 0, \quad F'(E^{1/N}) > 0 \quad \text{and} \quad F''(E^{1/N}) < 0. \]

The parameter \( \lambda \) describes how important the effort of the workers is for the firm. Low values of \( \lambda \) mean that production is very vulnerable if the workers reduce their flexibility, effort or care. Hence, in the firms with low \( \lambda \), the workers' commitment to proper working practices is more valuable than in the firms with a high \( \lambda \). In the latter group of firms, it is much easier to substitute labour for effort than in the firms with low \( \lambda \) and high \( \frac{1}{N} \).

Aggregate effort \( E \) is a function of all individual contributions. The following formulation is adopted:

\[ E(E_1, \ldots, E_N) = E_1^{\frac{1}{N}} E_2^{\frac{1}{N}} \ldots E_N^{\frac{1}{N}}, \]  \hspace{1cm} (3.8)

where \( N \) is the number of workers in the firm. The aggregate effort is assumed to be a symmetric Cobb-Douglas function\(^{20}\) of the individual efforts \( E_j, \ j = 1, \ldots, N \).

\(^{19}\)The operation of the basic model in its simplest form is analysed in Subsection 3.2.4.

\(^{20}\)The functional form, geometrical average, has a nice property

\[ A^\frac{1}{N} = A^{\frac{1}{N}} A^{\frac{1}{N}} = A^{\frac{1}{N}} A^\frac{1}{N} = \ldots = A. \]

An additional worker does not change the level of aggregate effort if he or she is of the same quality as the original workers. As a result, there is a clear trade-off between hiring more workers or paying more to the current ones to obtain a higher effort level. Straightforward calculations show that if \( N > 1 \), the \( E \)-function is increasing in \( E_i \) in a diminishing manner. As a consequence, there is no reason to use similar workers differently in production. See Propositions B.1 and B.2, page 155 onwards.
3.2. Basic Model

In what follows, it is assumed that all the workers are equal and that they are paid and used symmetrically in production. Then the production side of the model can be solved using aggregate variables. In Appendix B, it is shown that equal pay and equal role in production follow from equal quality of the workers.

The first step of the analysis is to derive the iso-profit curves. In what follows, the firms are assumed to be small and possess no power at the product markets, hence the price level can be treated as given. Profits $\Pi \equiv F(E^{1/\lambda}N) - WN$ are maximised when $\Pi'(N) = 0$; the optimality condition is

$$W = E^{1/\lambda}F'(E^{1/\lambda}N).$$

(3.9)

Because $F'(\cdot)$ is strictly decreasing, Equation (3.9) can be inverted and the following expression is obtained.

$$E^{1/\lambda}N = (F')^{-1}(WE^{-1/\lambda}).$$

(3.10)

Using one of the first-order conditions of the model allows us to write the argument of the production function in a more useful form.

The model is partly log-linear and it is, hence, practical to analyse it in the logarithmic form. Using logarithmic variables,

$$e \equiv \log(E)$$

(3.11)

and

$$w \equiv \log(W),$$

(3.12)

the variable

$$v \equiv e - \lambda w$$

(3.13)

can be defined. This variable will have a substantial role in the following analysis.

The next step in the analysis is to show that profit is an increasing function of the variable $v$. From Equation (3.13), one may note that

$$\exp(-v/\lambda) = WE^{-1/\lambda}.$$ 

(3.14)

Let us define the function $H_\lambda(v)$ as follows.

$$H_\lambda(v) \equiv (F')^{-1}(\exp(-v/\lambda)).$$

(3.15)

---

21Symmetric use implies that $W_1 = \ldots = W_N = W$ and $E_1 = \ldots = E_N = E$.

22Ramaswamy and Rowthorn (1991, 1993) is followed here. However, labour input is measured here in terms of the number of workers and wage is measured in per head terms. Ramaswamy and Rowthorn did not specify the unit labour input.
Using Equations (3.14) and (3.15), it is clear that the argument of the production function (see Equations (3.7) and (3.10)) is equal to the function $H_\lambda(v)$.

$$E^{\frac{1}{\lambda}} N = H_\lambda(v)$$  \hspace{1cm} (3.16)

As a by-product, labour demand can be solved.

$$N = H_\lambda(v)E^{-\frac{1}{\lambda}} = H_\lambda(v)\exp\left(-\frac{v}{\lambda}\right)$$  \hspace{1cm} (3.17)

Though the number of firms is constant, labour demand varies according to the profitability of each firm.

The wage bill $WN$ can be written in the form

$$WE^{-\frac{1}{\lambda}} \cdot E^{\frac{1}{\lambda}} N = \exp\left(-\frac{v}{\lambda}\right)H_\lambda(v).$$  \hspace{1cm} (3.18)

As a result, production and profit can be expressed as functions of the auxiliary variable $v$.

$$Y = F(H_\lambda(v))$$  \hspace{1cm} (3.19)

$$\Pi = F(H_\lambda(v)) - \exp\left(-\frac{v}{\lambda}\right)H_\lambda(v)$$  \hspace{1cm} (3.20)

Note that $F'(\cdot)$ is decreasing in $v$. Then also $(F'(\cdot))^{-1}$ is decreasing and as a result $H_\lambda(v)$ is increasing in $v$, see Equation (3.15). Because

$$\frac{d\Pi}{dv} = H_\lambda(v)\left[F'(H_\lambda(v)) - \exp\left(-\frac{v}{\lambda}\right)\right] + \frac{1}{\lambda}H_\lambda(v)\exp\left(-\frac{v}{\lambda}\right)$$  \hspace{1cm} (3.21)

and given that (see Equations (3.9) and (3.14))

$$\exp\left(-\frac{v}{\lambda}\right) = WE^{-\frac{1}{\lambda}} = F'(E^{\frac{1}{\lambda}} N),$$  \hspace{1cm} (3.22)

the following results are obtained.

$$\frac{dH_\lambda(v)}{dv} > 0, \hspace{0.5cm} \frac{dY}{dv} > 0 \hspace{0.5cm} \text{and} \hspace{0.5cm} \frac{d\Pi}{dv} > 0.$$  \hspace{1cm} (3.23)

In the optimum, production and profit of the firm with $\lambda$ depend solely on the variable $v$ and thus it is in the interest of the firm to maximise it. Iso-profit curves are therefore of the form $v = \text{constant}$. Using Equation (3.13) gives the form of the iso-profit curves in logarithmic efficiency and wage space.

$$e = \text{constant} + \lambda w$$  \hspace{1cm} (3.24)
It should be noted that the iso-profit curves are linear in logarithmic space for all concave production functions $F(\cdot)$ as far as $E^\frac{1}{N}$ is the argument of the function. This property is used, for example, in Figure 3.1 on page 59.

The firm seeks the highest effort and the lowest wage available. Combinations of $(E, W)$ or equivalently combinations of $(w, e)$ are, however, constrained by labour supply. Workers do not have any reason to accept high effort requirements if the offered wage is not high enough because an outside option is assumed to exist. Derivation of the effort function is the topic of the next subsection.

### 3.2.3 Preferences of workers

Until now, the effort function has been used without a formal derivation. The purpose of this subsection is to show that the concave effort function can be derived from a reasonable utility function.\(^{23}\)

The starting point of analysis of this chapter was that female and male workers are not equal. For a woman, it is more costly to commit oneself to work according to the rules of the employer. This can be modelled as follows. The utility of an individual worker is assumed to be an additively separable function of wage and effort performed.

$$U = U(W, E, \Phi) = \log(W) - V(\Phi E),$$

where $\Phi \geq 1.0,$

$$V'(\cdot) > 0, \quad V''(\cdot) > 0.$$  \(3.25\)

The wage $W$ enters into utility function logarithmically.\(^{24}\) The convex function $V$ describes the disutility resulting from the effort made. The parameter $\Phi$ is allowed to vary across individuals. The higher is $\Phi$, the higher is utility loss when effort is made.\(^{25}\)

By fixing utility to the level $\overline{U}$ and inverting the function $V$, effort $E$ can be written as a function of $W$, $\Phi$ and $\overline{U}$.

$$E = G(W, \Phi, \overline{U}) \equiv \frac{V^{-1} \left( \log(W) - \overline{U} \right)}{\Phi} \quad (3.26)$$

\(^{23}\)The utility function is given an alternative interpretation in Section 3.4.

\(^{24}\)The utility function used in this chapter is a simplified version of the more complex alternative, see Appendix B, page 157.

\(^{25}\)In basic analyses, $\Phi$ is taken as a given parameter describing preferences. In Subsection 3.4, it will have a wider interpretation. Moreover, determining its value is also discussed there.
A logarithmic version of (3.26) is

\[ e \equiv \log(E) = \log \left[ V^{-1} \left( w - \bar{U} \right) \right] - \phi \equiv g(w - \bar{U}) - \phi, \quad (3.27) \]

where \( \phi = \log(\Phi) \) and \( w = \log(W) \).

Because \( V(.) \) is increasing and convex, \( (V)^{-1}(.) \) is concave. The relation between logarithmic effort and logarithmic wage is thus also concave. This is shown in more detail in Appendix B (page 158), where partial derivatives of the \( g(.) \) function are also provided.

The link between individual effort levels and aggregate effort was defined in a way to allow easy aggregation\textsuperscript{26}. If the workers are identical, the aggregate effort function \( g(.) \) is also concave, i.e.

\[ e = g(w - \bar{U}) - \phi, \quad g'(w - \bar{U}) > 0, \quad g''(w - \bar{U}) < 0. \quad (3.28) \]

Linear iso-profit curves and convex iso-utility curves guarantee that the firm’s maximisation problem is well defined. The firm chooses the highest iso-profit contour consistent with the restriction (3.28). In Subsection 3.2.4 it will be shown that firms with different \( \lambda \)’s have optimums at different points of the \( g \)-curve.

### 3.2.4 Equilibrium with one gender and two firms

Using the tools developed above, wage distribution can be analysed. For simplicity, assume first that there is only one gender and two firms in the economy. The assumptions imply that \( \Phi_{ij} = \bar{\Phi} \) for \( i = 1, 2 \) and \( j = 1, \ldots, N_i \). 

\( N_i \) is number of workers in firm \( i \). Without loss of generality, we can set \( \bar{\Phi} = 1 \) and thus \( \phi_{ij} \equiv \log(\bar{\Phi}) = 0 \). There is therefore a single and simple aggregate effort function \( e = g(w - \bar{U}) \).

Let us begin the analysis by assuming that \( \bar{U} \) is constant and high enough to cause unemployment. The assumption will be relaxed later. Firms differ from each other only in how important effort is in production. The efficiency parameter for the Firm 1 is \( \lambda_1 \) and for the Firm 2 \( \lambda_2 \), \( \lambda_2 > \lambda_1 \). Because \( \frac{1}{\lambda_1} > \frac{1}{\lambda_2} \), effort is more important for Firm 1. The situation can be analysed using iso-profit lines

\[ e_1(w) = C_1 + \lambda_1 w \]

\[ e_2(w) = C_2 + \lambda_2 w, \quad (3.29) \]

where \( C_1 \) and \( C_2 \) are constants defining specific lines on the iso-profit curves.

\textsuperscript{26}See Appendix B, page 158.
Figure 3.1: Equilibrium with one gender and two firms

The firms choose the highest profit lines consistent with the $g$-function. They pick different points from the worker’s iso-utility curve $g(w)$ because the iso-profit lines are tangent to the $g$-function at different points, see Figure 3.1.

The easiest way to solve the problem is to maximise $C_i = e - \lambda_i w$ subject to $e = g(w - \bar{U})$. The restriction can be substituted in the maximand. The first order condition of the problem

$$
\max_w g(w - \bar{U}) - \lambda_i w
$$

(3.30)

is

$$
g'(w_i - \bar{U}) = \lambda_i
$$

(3.31)

and the second order condition is automatically satisfied because the $g$-function is concave\textsuperscript{27}. Because $g'(.)$ is monotonously decreasing (i.e. $g''(.) < 0$), it is clear that

$$
w_i = (g')^{-1} (\lambda_i) + \bar{U}
$$

(3.32)

\textsuperscript{27}If the function were linear, i.e. if the corresponding function between $E$ and $W$ were exponential ($E = \exp(a)W^b$) instead of concave, then there would be a corner solution.
and that
\[ e_i = g \left( (g')^{-1}(\lambda_i) \right), \]  
(3.33)
i.e. the effort taken is determined according to the production function and is independent of endogenous variables.

In the firm optimum, there are workers earning \( w_1 \) and producing effort \( e_1 \), workers earning \( w_2 \) and performing effort \( e_2 \) and some unemployed workers. This cannot be an equilibrium as the unemployed workers would start to compete for the jobs.

Using the labour demand equation (Equation (3.17)) one may obtain the labour market equilibrium condition
\[ N = \sum_{i=1}^{2} N_i = \sum_{i=1}^{2} H_{\lambda_i} \left( e_i - \lambda_i w_i \right) \exp\left( - \frac{e_i}{\lambda_i} \right), \]  
(3.34)
where \( N \) is the number of workers. Because \( e_i \) is constant and \( dw_i = dU \) by Equation (3.32), labour demand \( N_i \) depends negatively on the utility the workers enjoy, i.e.
\[ \frac{dN_i(U)}{dU} = -\lambda_i \frac{dH_{\lambda_i}(\cdot)}{dw} \exp\left( - \frac{e_i}{\lambda_i} \right) < 0. \]  
(3.35)
Equation (3.34) hence defines utility \( U \) uniquely; corresponding values of labour demands, wages and efforts can be calculated using it. As noted previously, the Firm 1 with low \( \lambda \) pays high wage and the workers in this firm work more effectively than those in Firm 2.

The results obtained can be generalised for more than two firms without difficulty. However, there is no reason to do this here as the results would be obvious; firms with high \( \lambda \) and thus low \( \lambda^{-1} \) find their optimum at a low wage and effort level. These firms value effort low and are capable of substituting labour for effort and care.\(^{28}\)

---

\(^{28}\)An example of this kind of activity is cleaning work. It is always more profitable to hire more workers than to pay more to a constant number of staff. An opposite example is airline piloting. To fulfill the task of safe navigation it is better to pay more than to employ more people into the cockpit of the aeroplane.
3.3 Heterogeneous workers

In the previous section, the basic model was set up and analysed. In the first phase a production function with two labour inputs, effort and the number of workers, was postulated and the shape of corresponding iso-profit contours was derived. The second step was to derive the effort function, i.e. iso-utility curve for the workers from the utility function. In the following subsection the setting was completed by an analysis of market equilibrium when there is only one type of worker. If there are no gender differences in the labour force, then those firms with high effort requirements pay higher wages than are paid by the firms with low-effort production technologies.

In this section, gender dissimilarities discussed at Section 3.1 are introduced into the formal analysis. In the model, the genders differ from each other with respect to the cost of effort, in all other ways they are equal. This assumption is based on the discussion in Subsections 3.1.1 and 3.1.2 where it was argued that taking care of children requires mainly effort, but not so much labour measured in terms of hours. Hence, women, who are usually in charge of taking care of children, find it harder than men to commit to work efficiently, i.e. under the strict rules of an employer. High effort can be interpreted in this context as workers’ readiness to concentrate on their duties, to work flexible hours and to leave personal matters out of their mind when working.

The section is organised as follows. The analysis starts with the case in which society is divided into two genders. The two genders are internally homogeneous, but differ from each other. Basic properties of the equilibrium are analysed in Subsection 3.3.1 using the notation of the previous section. The results are illustrated graphically in Appendix B. A continuous firm distribution is assumed as it makes our notation much more simple and the results more transparent. The main result of the analysis is presented in Subsection 3.3.2.

The assumption that the genders are internally homogeneous is an extreme one. Subsection 3.3.3 deals with a more realistic case where individual workers are allowed to differ from each other even when they are of the same gender.
3.3.1 Equilibrium with two genders - discrete case

Gender differences can be modelled by having different $\Phi$ for male workers and for female workers. For simplicity, we assume that $\Phi_M = 1.0$ for male and $\Phi_F > 1.0$ for female workers. The corresponding logarithmic coefficients are $\phi_m \equiv \log(\Phi_M) = 0$ and $\phi_f \equiv \log(\Phi_F) > 0$. In what follows, the subscript $f$ is dropped from $\phi_f$ because the symbol $\phi_m$ is not in use as $\phi_m \equiv 0$.

The effort function $e = g(w - U)$ defines the shape of the iso-utility curvature. By varying $U$, one may observe how the shape of the iso-utility contour changes with utility levels. From Equations (B.8) and (B.10)\(^{29}\) it can be noted that the iso-utility curves for women are of the same shape as those for men, but are located some distance $\phi$ below the corresponding male curves in logarithmic wage and efficiency space. Hence a female worker’s effort is lower than the effort of a male worker at the same utility level. This creates a preference for the firms; male workers are preferred to female workers. However, market mechanisms correct the situation and the utilities of the genders adjust. Let us denote male utility as $\bar{U}$ and female utility as $\bar{\bar{U}}$.

**Proposition 3.1** In an equilibrium, the utility of male workers exceeds the utility of female workers, i.e. $\bar{U} > \bar{\bar{U}}$.

**Proof** We assume that $\bar{U} \geq \bar{\bar{U}}$ and show that it impossible. A firm may hire a male worker with a $(e, w)$ combination determined by the function $e = g(w - \bar{U})$ or a female worker with $(e, w)$ combination determined by function $e = g(w - \bar{\bar{U}}) - \phi$. Because $\bar{U} \geq \bar{\bar{U}}$, $\phi$ is positive and function $g(.)$ is increasing, $g(w - \bar{U}) > g(w - \bar{\bar{U}}) - \phi$ for every $w$. This cannot be an equilibrium allocation since every firm with a female worker is willing to replace her with a man and every man is ready to accept the replacement as his utility does not decrease. \(\square\)

Consider next the market allocation of genders.

**Assumption 3.1** There are $h$ firms in the economy. The firms are characterised by their effort parameters $\lambda_i$, $i = 1, \ldots, h$. Let us rank the firms according to the size of the parameter:

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{h-1} \leq \lambda_h.$$  

\(^{29}\)See Appendix B, page 158.
Firm number 1 values effort at least as much as Firm 2 etc. Intuitively it is clear that male workers are concentrated in firms with low serial numbers. Let us analyse the problem more carefully. The following Assumption 3.2 simplifies the analysis of gender allocation without loss of generality.

**Assumption 3.2** All workers in one particular firm are of the same gender.

**Proposition 3.2** Let Assumptions 3.1 and 3.2 hold. In a market equilibrium it is impossible that women are working in the firm with $\lambda_i$ and men in the firm with $\lambda_j$ if $\lambda_i < \lambda_j$.

**Proof** See Appendix B.

Proposition 3.2 has a clear interpretation. Female workers are concentrated in firms with a high $\lambda$, i.e. with low effort requirements. Male workers are employed by firms with low $\lambda$ and high $\frac{1}{\lambda}$. This result is illustrated graphically in Appendix B. The next step is to study the determination of the utilities $\bar{U}$ and $\bar{U}$. In what follows, to simplify the analysis, an assumption on continuous firm distribution is adopted.

### 3.3.2 Equilibrium with two genders - continuous case

Consider an allocation of workers in which men are concentrated in the firms $i = 1, \ldots, s$ and women in the firms $i = s + 1, \ldots, b$. Denote number of male and female workers in the economy by $N_m$ and $N_f$. Gender utilities $\bar{U}$ and $\bar{U}$ as well as the parameter $s$ could be solved using the definition of the iso-profit curve and labour market equilibrium conditions

$$N_m = \sum_{i=1}^{s} N_i^m(\bar{U}) \quad \text{and} \quad N_f = \sum_{i=s+1}^{b} N_i^f(\bar{U}), \quad (3.36)$$

where $N_i^m(\bar{U})$ is labour demand in a firm employing only men and $N_i^f(\bar{U})$ labour demand in a firm employing only women. However, it turns out that the analysis is more transparent and the notation simpler if a continuous case is analysed, but the results are not altered.

**Assumption 3.3** Firms are indexed according to the efficiency parameter $\lambda$, which is uniformly distributed between $\underline{\lambda}$ and $\overline{\lambda}$. To avoid expressing the density function explicitly, assume that $\overline{\lambda} - \underline{\lambda} = 1$. 
Utilities $\bar{U}$ and $\bar{U}$ and the parameter $s$ can be solved from the following equilibrium conditions.

$$N_f = \int_s^{\bar{N}} N_i^f (U) \, di$$  \hspace{1cm} (3.37)

$$N_m = \int_s^{\bar{N}} N_i^m (U) \, di$$  \hspace{1cm} (3.38)

$$s = \frac{\phi}{\bar{U} - \bar{U}}$$  \hspace{1cm} (3.39)

Start with the female labour market equilibrium condition (Equation (3.37)). Firms from $s$ to $\bar{N}$ employ women; the labour demand of an individual firm $i$ is

$$H_{\lambda_i} \left( g(w_i - \bar{U}) - \phi - \lambda_i w \right) \exp \left( - \frac{g(w_i - \bar{U}) - \phi}{\lambda_i} \right).$$  \hspace{1cm} (3.40)

The value of $g(w_i - \bar{U})$ and thus effort $e_i$ are given because $\lambda_i = g'(\cdot)$. Because $H_{\lambda_i} (\cdot)$ is increasing and $dw_i = d\bar{U}$ (see Equation (3.32)), labour demand $N_i$ depends negatively on the utility the workers enjoy. Labour market equilibrium requires that any decrease in labour demand in the firms using female labour is absorbed by the other firms. Hence, the higher utility, the lower the parameter $s$. The equilibrium condition (Equation (3.37)) can be written in the form

$$\bar{U} = U(s), \quad \bar{U}'(s) < 0.$$  \hspace{1cm} (3.41)

By similar reasoning one obtains

$$\bar{U} = \bar{U}(s), \quad \bar{U}'(s) > 0.$$  \hspace{1cm} (3.42)

The third equilibrium condition (Equation (3.39)) is based on linearity of the iso-profit curves $v$. In the equilibrium, firm $s$ should be indifferent about using female or male workers. Hence

$$v_f \equiv g(w_f - \bar{U}) - \phi - sw_f = g(w_m - \bar{U}) - sw_m \equiv v_m.$$  \hspace{1cm} (3.43)

Profit maximisation requires that

$$\frac{dg(w_f - \bar{U})}{dw} = \frac{dg(w_m - \bar{U})}{dw} = s,$$  \hspace{1cm} (3.44)
hence
\[ g(w_f - \bar{U}) = g(w_m - \bar{U}) \]  \hspace{1cm} (3.45)
and
\[ w_f - \bar{U} = w_m - \bar{U}. \]  \hspace{1cm} (3.46)
From these equations one obtains
\[ s = \frac{\phi}{w_m - w_f} = \frac{\phi}{\bar{U} - \bar{U}}, \]  \hspace{1cm} (3.47)
i.e. Equation (3.39). Substituting Equations (3.41) and (3.42) into Equation (3.47) gives
\[ s = \frac{\phi}{U(s) - \bar{U}(s)}, \]  \hspace{1cm} (3.48)
which combines three equilibrium conditions in a neat form. As the right-hand-side of (3.48) is decreasing and the left-hand-side is increasing in \( s \), there is a unique solution \( s^* \) and the following conclusion is justified. It is the main proposition of this chapter.

**Proposition 3.3** Let Assumptions 3.1 - 3.3 hold. The gender gap, i.e. utility difference between the genders is
\[ \bar{U} - \bar{U} = \frac{\phi}{s^*}, \]
where \( \phi \) describes the extra effort cost of women and \( s^* \) is a parameter describing the importance of effort in the firm which is indifferent about employing men and women.

The wage premium of the model should be interpreted carefully. It is the difference between male and female wages in firm \( s^* \), i.e. in the firm which is indifferent between male and female workers. At the same time, the premium is the difference between the highest female wage and the lowest male wage. The wage paid to male workers deviates from the female wage more when the importance of effort increases in firm \( s^* \). The same factor determines the extra utility, a sort of a "gender rent", that male workers enjoy. In that sense, firms at the "middle" of the firm distribution are crucial in determining gender stratification in the labour market.
3.3.3 Varying family commitments

The analysis of Subsection 3.3.2 was based on an assumption that female and male workers had different family commitments. Also, it was assumed that these commitments are equal within both genders groups. As a result, it was found that the assumed higher female cost of labour market commitments led to a polarised labour market. Male workers are employed in high effort jobs and are paid well, and female workers are working in low paid, not particularly demanding professions. As a consequence of our extreme assumptions, the model predicted that the highest female wage (effort) is lower than the lowest male wage (effort). In this subsection, we show that if this unrealistic assumption is relaxed, then the unrealistic predictions also vanish.

As earlier, an economy with \( h \) firms, \( N_f \) female workers and \( N_m \) male workers is studied and Assumptions 3.1 - 3.3 hold. On the workers’ side, there is now more variation than before. Instead of two different values, \( \Phi \) may take any value larger than or equal to one. Hence, the logarithmic \( \phi_i \)’s \( (i \) is index of individual, \( i = 1, \ldots, N_f + N_m) \) are positive. Without loss of generality individuals may be ranked according to their \( \phi_i \)’s as follows.

\[
\phi_1 \leq \phi_2 \leq \ldots \leq \phi_{N_f+N_m} \quad (3.49)
\]

**Assumption 3.4** There are \( k \) \( (k \leq N_f + N_m) \) groups of workers such that for every worker belonging to group \( i, i = 1, \ldots, k \) there is an effort parameter \( \phi_i, \phi_1 < \phi_2 < \ldots < \phi_k \). In other words, the groups are internally homogeneous. Let us denote the size of groups by \( N_i \).

Immediately one may conclude that all members in group \( i \) have the same utility \( U_i \). Because of asymmetrical gender roles it may be assumed that male workers are concentrated in the beginning of the sequence and female workers in the end, but in the middle they are mixed. However natural, this assumption is not necessary.

Utilities \( U_i, i = 1, \ldots, k \) and parameters \( s_1, s_2, \ldots, s_{k-1} \) are determined according to the following equilibrium conditions.

\[
N_k = \int_{s_{k-1}}^{\infty} N_i(U_k) \, di \quad (3.50)
\]

\[
N_{k-1} = \int_{s_{k-2}}^{s_{k-1}} N_i(U_{k-1}) \, di \quad (3.51)
\]
\[N_1 = \int_0^{s_1} N_i(U_1) \, di\]  
(3.52)

\[s_{k-1} = \frac{\phi_k - \phi_{k-1}}{U_{k-1} - U_k}\]  
(3.53)

\[s_{k-2} = \frac{\phi_{k-1} - \phi_{k-2}}{U_{k-2} - U_{k-1}}\]  
(3.54)

\[s_1 = \frac{\phi_2 - \phi_1}{U_1 - U_2}\]  
(3.55)

In the system, there are \(2k - 1\) unknown variables and \(2k - 1\) equations, hence it can be solved. Because

\[U_1 - U_2 = \frac{\phi_2 - \phi_1}{s_1} \quad \text{and} \quad U_2 - U_3 = \frac{\phi_3 - \phi_2}{s_2},\]  
(3.56)

Proposition 3.3 can be generalised.

**Proposition 3.4** Let Assumptions 3.1 - 3.4 hold. The utilities of \(k\) groups of workers are

\[U_1 = U_2 + \frac{\phi_2 - \phi_1}{s_1}\]

\[= U_3 + \frac{\phi_3 - \phi_2}{s_2} + \frac{\phi_2 - \phi_1}{s_1}\]

\[\vdots\]

\[= U_k + \frac{\phi_k - \phi_{k-1}}{s_{k-1}} + \ldots + \frac{\phi_2 - \phi_1}{s_1},\]

where \(\phi_i\)’s are group-specific effort cost parameters and \(s_i\)’s effort parameters of the firms that are indifferent about employing workers from groups \(i\) and \(i + 1\). Differences \(U_i - U_{i+1} = (\phi_{i+1} - \phi_i)/s_i\) are positive.

It can be concluded that the model is capable of producing a wage and utility distribution. If men generally have less childcare responsibilities than women, they are more often than women allocated to jobs with a high wage and utility.
3.4 Modernisation of family roles

In the previous sections family roles were analysed as exogenous for the families. In the section at hand, family roles are analysed as endogenous choices. Every family must provide a minimum amount of care for the children and there is a trade-off between the efforts of the parents. A family decides on its gender roles taking into account the expected wage distribution and the labour market location of the family members.

In the basic model, the efficiency parameter $\Phi$ was interpreted as a parameter describing preferences. In what follows, the preference interpretation will be relaxed and the parameter is merely understood as an outcome from underlying preferences and restrictions, affected by public policies.

For technical reasons, it is assumed that the family cannot freely choose the roles because this kind of model would be too complex to solve. In the model, there is only one alternative for the traditional family model. A modern family pattern\footnote{An emergence of new family roles may be a result of the invention of new childcare technologies. For example, in a traditional society, a change may be caused by the availability of bottled milk. Children could then be cared for by their father as well as their mother. Another example is the emergence of egalitarian values. Families may recognise that the responsibility for children can be organised in a new way.} is adopted if it renders higher utility than the traditional one. If new roles are adopted, more male effort is required but some additional female energy is saved for the labour market. A perfectly egalitarian assumption is adopted.

**Assumption 3.5** If a family adopts a modern family pattern, the load of childcare is divided exactly evenly.

Formally, this can be modelled as follows. A family, i.e. a couple consisting of a woman and a man, has collective childcare duty. Either the female member of the family bears the whole load, i.e. childcare parameters are $\Phi_f$ for women and $\Phi_m = 1$ for men or the load is divided such that $\Phi_n$ is parameter for both members of the family. The corresponding logarithmic parameters are $\phi_f, 0$ and $\phi_n, \phi_m$. The corresponding utilities are $U, \bar{U}$ (for the traditional women and men) and $U_n, U_n$ (for the members of the modern families).

**Assumption 3.6** The following holds for the parameters: $\Phi_f > \Phi_n > 1$. 
Proposition 3.5 Let Assumption 3.6 hold. Then $\bar{U} > U_n > \bar{U}$.

Proof Assumption 3.6 and Proposition 3.4 together imply that $\bar{U} > U_n > \bar{U}$. $\square$

The family faces a choice that will affect family life as well as labour market participation of the family. According to Proposition 3.5, the following proposition is also true.

Proposition 3.6 Deviation from the traditional family roles increases the utility of the female member of the family and decreases the utility of the male member.

Without altruism, centralised decision making in the family or side payments, any deviation from the traditional family pattern is therefore unlikely; Pareto improvements are impossible. Altruistic motives and side payments may change the situation. However, the childcare technology described by parameter pairs $\phi_f, 0$ and $\phi_n, \phi_n$, should favour division of duties. That is, $\phi_n$ should be small relative to $\phi_f$. Only then the change in the utility of the female worker can be large relative to the change in the utility of the male worker, i.e. $U_n - \bar{U}$ should be large relative to $\bar{U} - U_n$. The situation can be modelled as follows.

The parameter $\Phi$ was introduced into the utility function (Equation (3.25), page 57) as a parameter describing individual preferences. If Equation (3.25) is interpreted as a derivative of underlying preferences and constraints, then $\Phi$ also might have a wider interpretation.

Assumption 3.7 Childcare is characterised by a Cobb-Douglas household production function

$$\Phi_f^\phi \Phi_m^\beta \Phi_g^\gamma \geq \Phi \gg 1,$$

where minimum childcare $\Phi$ is produced by the efforts of the wife $\Phi_f$, the husband $\Phi_m$ and the government $\Phi_g$. $\Phi$ is high enough to guarantee that there are still some responsibilities for the parents after any government intervention. Some basic effort is required from everyone involved in the production, let us denote this by $\Phi_f \geq 1$, $\Phi_m \geq 1$ and $\Phi_g \geq 1$.

In the basic two-gender case (see Subsections 3.3.1 - 3.3.2) we assumed implicitly that $\Phi_m = \Phi_g = 1$, i.e. that the childcare loads of men and the government are minimal. Because effort is arduous, women minimise it and set

$$\Phi_f = \Phi_m^{1/\alpha} = \Phi.$$

(3.57)
Let Assumption 3.7 hold and express the childcare production technology in logarithmic form.

\[ \alpha \phi_f + \beta \phi_m = \tilde{\phi} - \gamma \phi_g \equiv \hat{\phi}, \]  

(3.58)

where \( \phi_f, \phi_m \) and \( \phi_g \) are childcare loads of the wife, the husband and the government. \( \alpha, \beta \) and \( \gamma \) are household production function parameters and \( \hat{\phi} \) the minimum care requirement. If childcare is solely a female duty, i.e. if \( \phi_m = 0 \), then

\[ \phi_f = \frac{\hat{\phi}}{\alpha}. \]  

(3.59)

In the modern families, the load is divided equally, i.e. \( \phi_f = \phi_m = \phi_n \) such that

\[ \phi_n = \frac{\hat{\phi}}{\alpha + \beta}. \]  

(3.60)

**Assumption 3.8** Families are utilitarian, i.e. families tend to maximise the sum of the female and the male utility.

As a direct consequence of Assumption 3.8, one may observe that the following must hold in the market equilibrium:

(i) \[ U + \bar{U} \geq 2U_n \] if all the families are traditional,

(ii) \[ U + \bar{U} = 2U_n \] if both family types exist, and

(iii) \[ U + \bar{U} \leq 2U_n \] if all the families are modern.

A stationary situation prevails only if there is no way to gain by changing the roles in the family.

**Proposition 3.7** Let Assumptions 3.5 - 3.8 hold.

(a) There are modern families only if men are at least as productive in childcare as women, i.e. \( \beta \geq \alpha \).

(b) If men and women are equally productive in childcare, i.e. if \( \beta \) approaches \( \alpha \), the number of modern families approaches zero.

(c) The share of modern families approaches one only if men are remarkably more productive in childcare than women, i.e. only if

\[ \frac{\alpha + \beta}{\alpha} \geq \frac{\lambda + \lambda}{\lambda}, \]

i.e. only if \( \beta \gg \alpha \).

(d) Government policy does not affect family decision making.
**Proof** See Appendix B.

Proposition 3.7 may seem quite pessimistic in terms of egalitarian aims. Families adopt equal roles only if men are at least effective in childcare than women. However, one may ask why families would then adopt an equal division of the childcare load. Why would the most productive member of the family, the man, not specialise in childcare?

The results obtained depend, of course, heavily on the strong assumptions made. However, they highlight some general features of our model and, perhaps also of the real world. First, when a group of individuals with equal effort parameters becomes larger, members of the group face declining utility. Second, in a model where gender stratification arises from exogenous differences between the sexes, all gender gaps remain until the basic factors are changed. In this respect, our results are more pessimistic that those of Francois (1998). In his analysis, there are no exogenous gender differences, therefore relatively weak policy interventions can make a difference. For example, an affirmative action programme would be enough to shift families from unequal gender roles to equal ones. In the analysis at hand, this would not be enough as the differences are more deeply rooted in societies and cultures. There is therefore no reason to expect very good output-to-input ratio for policies targeting more equal labour market stratification.
3.5 Concluding remarks

3.5.1 Predictions of the model

We have analysed labour market equilibrium in the case of unequal childcare loads. According to the theoretical analysis, those with heavy domestic commitments, or women, are allocated to occupations with predictable timetables and without high commitment requirements, i.e. in time-intensive low-pay production processes. Men, with lighter domestic commitments, concentrate instead in firms where effort, i.e. commitment of the employees, is more important. Flexible and irregular hours and a high commitment to serve the employer are typical in, let us say, managerial professions. Moreover, it was predicted that firms pay higher wages to workers who are committed to work effectively and flexibly, thus women earn less than men.

Our second major conclusion was concerned with the size of the wage difference, i.e. the wage gap. According to our simplified model, the difference between childcare loads of the genders (parameter $\phi$ in the model) explains wage gaps. The higher the difference in load, the higher wage gap. The number of children is, of course, an important determinant of the load, but also the childcare facilities available and the division of duties within family matter. Keeping industrial societies in mind, it was argued that the availability of public childcare facilities, e.g. daycare centres, might have an effect on wage gaps.

3.5.2 Does the theory fit with the facts?

There are many factors affecting the gender wage differential: traditions (i.e. inertia in social roles and practices), differences in skills and education and, clearly, discrimination. The theory presented in this chapter does not aim to substitute all other explanations, it is merely a complement. However, for analytical and expositonal reasons, the theory is presented using a simplified model without any traditional elements creating gender wage differentials. For the same reasons, the results obtained should not be taken too literally. It is, however, interesting to note that in the two-gender case, an extremely small difference in domestic effort requirements causes a totally diversified gender allocation in the labour market. Though this is
a property of our simplified model, it might be a sign of something more
general. Like Becker (1993) points out, the cause and the consequences may
be extremely disproportionate in some cases. Hence, calculations showing
only a small productivity difference between genders may be misleading;
it is possible that minor differences have major consequences. If this is
the case, empirical testing of the theory presented here would be quite a
arduous task.

Before more structured analyses are available, one may consider some spe-
cial cases and casual observations. Our analysis of wage dispersion was
based on the efficiency wage model by Ramaswamy and Rowthorn (1991,
1993). They used airline piloting and cleaning as examples of the main idea;
persons with high effort requirements are paid high wages as a compensa-
tion for effort produced. Pilots are required to work in a very concentrated
manner; lack of effort cannot be compensated by longer working hours. On
the other hand, a bad cleaning result can be corrected by appointing more
cleaners or doing the job several times. The same example is suitable for the
analysis of gender wage differential. Almost all pilots are men and almost
all cleaners are women.

Considering the care professions complicates and enriches the picture.
For example, both doctors and nurses should work carefully. Also, the work
process is continuous in both professions. Patients cannot be left alone even
if the family commitments of the staff would require one to stop working.
Still there is a big difference in the salaries of doctors and nurses. Though
this observation at first glance seems to contradict the theory presented
here, this is not the case. Not only the allocation of female workers adjusts,
but also the professional practices change according to the availability of
labour.

Care was historically organised at home and it was natural for women to
continue practising it when it was professionalised. Many researchers have
indeed emphasised that female wage labour emerged as an extension of
household work. Even nowadays, food processing, textile work and care are
performed by women both at home and at work. This is possible because
work processes are organised in a way that facilitates the female dual role.
For example, the shift system at hospitals is different for nurses and doctors.
Doctors, who are mostly male, are committed to work flexible hours, but
nurses, normally female, are working according to a rigid timetable. One
explanation is the fact that the fixed hours system suits also married women.
Moreover, hospitals offer nurseries for the children of the staff even in such
countries where there are generally no such services available.\textsuperscript{31}

The second theme of the analysis was the determinants of childcare loads. We argued that public childcare arrangements may be influential explanatory variable for the wage gaps. It is interesting to note that wage gaps\textsuperscript{32} are smallest in those European countries with large childcare programmes, see Table 3.1. Moreover, wage gaps in USA, Japan and Canada are remarkably higher than in Europe (see e.g. Gunderson 1989)\textsuperscript{33}. One possible explanation for this observation is the almost complete non-existence of public childcare policies in those countries. However, it cannot be emphasised too much that the evidence is very weak.

Waldfogel's (1998) empirical results are also interesting. According to her, there is a gap of 10-15 per cent between earnings of the women with children and of those without children. Women with children are systematically paid a lower wage than women without children after adjusting for differences in human capital attributes. On the other hand, married men receive a wage premium. Waldfogel includes short maternity leaves and insufficient childcare policies on the list of possible explanations for the findings.

Though casual observations support the theory, the acceptance of the theory, i.e. acceptance that the family commitment is a factor among the others affecting labour market stratification and that the availability of childcare may matter, would require systematic empirical research. Unfortunately this is not yet available. The following questions deserve attention in the future.

\textsuperscript{31} Not all women marry or have children. However, from the employers' side there is always a risk of pregnancy and a sudden increase in domestic duties. To diminish the risk, marriage has been forbidden for women in many professions. Maids, air hostesses and nurses have had to commit not to marry and the practice still prevails in some societies.

\textsuperscript{32} Duncan (1996) measures the wage gap by calculating the ratio of average female wage rate to average male wage rate. The ratio measures inequality in wage rates per unit time worked, not income differences. The figures should be interpreted very cautiously as the effects of the other factors affecting wage differentials, e.g. educational differences, are not controlled.

\textsuperscript{33} O'Neil and Polachek (1993) show, however, that the wage gap has narrowed during the 1980's in the US labour markets. On average, the gender gap declined by about 1% per year. The authors find that convergence in measurable work-related characteristics explains one-third to one-half the narrowing.
### Table 3.1: The wage gap: Western Europe around 1990 (Duncan 1996)

<table>
<thead>
<tr>
<th>Country</th>
<th>Wage gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>90</td>
</tr>
<tr>
<td>Denmark</td>
<td>87</td>
</tr>
<tr>
<td>Norway</td>
<td>86</td>
</tr>
<tr>
<td>Italy</td>
<td>83</td>
</tr>
<tr>
<td>France</td>
<td>81</td>
</tr>
<tr>
<td>Finland</td>
<td>78</td>
</tr>
<tr>
<td>Belgium</td>
<td>76</td>
</tr>
<tr>
<td>Greece</td>
<td>76</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>75</td>
</tr>
<tr>
<td>West Germany</td>
<td>73</td>
</tr>
<tr>
<td>Spain</td>
<td>72</td>
</tr>
<tr>
<td>Portugal</td>
<td>72</td>
</tr>
<tr>
<td>Ireland</td>
<td>68</td>
</tr>
<tr>
<td>UK</td>
<td>68</td>
</tr>
</tbody>
</table>

In the discussion, the terms "effort" and "flexibility" were used as synonyms. How valid is this interpretation? Is the distinction between effort and labour input measured in hours a useful conceptualisation? Can effort be measured? If yes, is there a positive correlation between the effort performed and the wage level? Is there a negative correlation between effort requirements and the share of women in different professions or in different firms? Is it possible to compare quantitatively childcare loads in different countries?

A study of the strength of commitments in typical female and typical male professions would be a reasonable starting point for combined conceptual and empirical research. International comparisons of childcare policies also deserve more attention by economists.

### 3.5.3 Is there any scope for change?

A hypothetical case with two alternative family roles, i.e. traditional and modern, was analysed in Section 3.4. The modern family was one with equal childcare commitments. It was shown that a part of the families adopt the modern family pattern only if the men are more productive in
childcare production than women. Otherwise, the modern option would imply in those families larger losses for the husband than the wife would gain. In this respect, our results were much more pessimistic than those of François (1998). According to his analysis, the gender roles are endogenous, and limited policy action would be enough to move the economy into a non-discriminating equilibrium.
Chapter 4

Finnish wage arbitration - theoretical comments

4.1 Introduction

The aim of this chapter is to analyse the operation and the role of the Finnish wage arbitration system; this hasn’t been attempted before in economic literature. However, easy answers are not available as the system is difficult to interpret. In this sense, the situation resembles its American and Canadian counterparts. In spite of large and growing economic literature on the topic\(^1\), many issues and aspects of arbitration are unclear and difficult to interpret. The analysis of this study is carried out partly by comparing Finnish and North American systems, partly by using simple game and decision theoretical arguments.

The correspondence between the Finnish arbitration system and the arbitration in the North American public sectors is not a well-established fact. As far I know, the link has not been studied before\(^2\). Hence, one of the first steps in the analysis is to show that the arbitrator may affect the wage outcome also in the Finnish system; in the North American system this is evident. After that, the issues raised in the North American discussion are analysed in the Finnish context.

The most important disagreement in the discussion has been about the

---


\(^2\)Professor Matti Pohjola suggested to me to study the topic. He is perhaps the first person to discover the link.
general role of the arbitrator in wage setting. As Bloom and Cavanagh (1987, page 357) put it:

... it remains an open empirical question whether arbitrators should mainly be viewed as individuals who (1) impose on negotiators their exogenous preferences, or (2) seek to learn about disputants’ preferences from the relationship between the facts and final positions in an attempt to search for outcomes that maximise the disputants’ welfare.

Kennan and Wilson (1993) strongly favour the first alternative, the second one is supported, for example, by Farber and Bazerman (1986).

The chapter is organised as follows. In the next two sections the Finnish arbitration system is described and North American arbitration systems, both existing and suggested ones, are discussed. Sections 4.4 and 4.5 discuss the theoretical fundamentals for understanding arbitration. It is argued that implementing desirable wage functions requires complicated structures. From that basis, the Finnish wage arbitration system is analysed with simple game theoretical tools in Sections 4.6 and 4.7. It is argued that the Finnish system can be interpreted as a particular form of arbitration, i.e. that it goes further than simple mediation or conciliation. Though there is a right to strike, the whole arbitration process is highly structured and the state arbitrator is very influential. Section 4.8 concludes the analysis.
4.2 The Finnish arbitration system

According to Finnish labour law, all strikes or lock-outs must be announced to the State Arbitrator\(^3\), hereafter the arbitrator, fourteen days before the industrial action is to take place. Otherwise, the action is illegal and those responsible can be prosecuted. Illegal strikes or lock-outs have been very uncommon recently and the rules of the arbitration system are well-honoured also in practice. There are several reasons for this. First, Finnish industrial relations have been generally good, at least in a comparative sense, for decades and strikes as such have been quite rare during the 1980’s and the 1990’s. Second, the state has historically been active in a tripartite system in the labour market. Incomes policy and nation-wide wage agreements have been important elements in the national consensus. Hence, it was easy for the trade unions and the employer’s associations to accept state involvement in the wage negotiation process; it was only one form of the state intervention in the labour market.

It is very difficult to believe that the current system of wage arbitration would have survived for more than thirty years without sincere support of the employers’ associations and the trade unions. I suggest that the arbitration system must be implicitly agreed upon by the labour market organisations. Although punishments for illegal industrial actions are severe, it is impossible to imagine a situation where all trade unions, firms or employers’ associations are forced to submit all cases to official arbitration contrary to their own will\(^4\).

4.2.1 Historical background

Conciliation and mediation of industrial disputes have a long history also in Finland\(^5\). Between the First and the Second World War, conciliation

---

\(^3\) **Vaikutusmallivirkistely** in Finnish. Kauppinen (1988) uses an alternative translation **The State Conciliator**. Although that is the word-for-word translation, I prefer to use the term **arbitrator**. I like to emphasis the fact that the Finnish arbitration system is highly structured and the arbitrator is influential. However, the term **conciliation** is proper when the period prior the Second World War or the first two decades after the war are described.

\(^4\) The very fact that the late State Arbitrator, Mr. Jorma Reini was chairperson of one of the Central Trade Unions before his nomination, supports the same conclusion. It also highlights the tripartite nature of the Finnish employment policy. Several other arbitrators have also had personal relations with the tripartite incomes policy.

\(^5\) Mattila (1992) mentions that the first known conciliation case was the strike in 1491 in the harbour of Turku. Mediation and conciliation became more important after the
was organised by the district conciliators appointed by the Ministry of Social Affairs (Mattila 1992). During the Second World War labour market disputes were mediated by wage control authorities. Wages were regulated until 1955, but regulation could not always cope with pressure from the trade union side. In fact, incomes policy with tripartite negotiations started during the 1950’s within the wage control councils. These councils were nominated to implement the laws defining wage increase rules. Trade unions and the employers’ associations had their own representatives in the wage councils. The role of these councils changed gradually; control was replaced by conciliation and agreements. (Sovala 1991a and 1991b; see also Pekkarinen and Vartiainen 1993)

After the Second World War, industrial relations were poor and conciliation was needed more than before. An act passed in 1946 by the Parliament made it obligatory for the disputing parties to give advance notice of a strike or a lock-out to the district conciliators and to the opposite party in the negotiations. Disputing parties were also obligated to take part in conciliation. (Kauppinen 1988) Conciliation started gradually to transform to arbitration.

When wage control was abolished in 1955, new institutional arrangements were needed. The main long term intention of the government was to create an incomes policy; such a policy finally emerged in the late 1960’s. A more modest aim was to reform wage conciliation. The reform started in 1955 and the law was passed in the Parliament in 1962. The main element of the law was the establishment of the independent office for the arbitrator. However, it is important to note that the arbitrator is a civil servant and is nominated for a four year term by the President of the Republic of Finland. Five district arbitrators are also nominated to assist the arbitrator.

---

6 Wage control was based on an idea of a constant consumption wage, i.e. all productivity gain was assumed to belong to the employer. However, the development of the labour market and the rapid economic growth that was experienced after the war were not consistent with conservative wage increases. Wage control was under constant pressure and wages actually increased faster than law allowed for.

7 Left-wing parties close to the trade unions voted against the reform in the Parliament. It was not an objection to the arbitration system as such, but merely a protest against some details of the system. For example, the Ministry of Social Affairs was given by the same law a right to postpone any strike or lock-out for two weeks and the trade unions did not want accept that.
4.2.2 Wage setting and arbitration in Finland

Wages are mainly agreed upon at the industry level in Finland. Since the 1960’s these agreements between trade unions and employers’ associations have been co-ordinated at the national level. Almost every year there has been at least an attempt to reach a centralised agreement between the central organisations of trade unions and the employers’ national bodies. During some years, also the State and the federation of farmers have joined in the agreement. The centralised agreement is binding only if the parties at the industry level agree to abide by it. In that sense, there are two negotiation rounds. The first round is about the content of the agreement and the second about who participates in the agreement.

Labour market conflicts in the Finnish system can be classified into three different types; the arbitrator is not involved in all of these. The first type is when there are problems in reaching a central wage agreement at the national level. These disputes are normally mediated by a top civil servant or by a leading politician, e.g. the Prime Minister, and the arbitrator is very seldom involved. Because reaching centralised agreement is not obligatory in the Finnish system, these disputes normally end without a deal. The parties then engage in industry-level negotiations, in which agreements are reached without any formal co-ordination.

The second type of conflict takes place when there is disagreement at the industry level. These disputes are the main field of work of the arbitrator. He gets involved in the process after one of the parties has filed a threat of industrial action.

The third type is disputes at the firm level. These may happen if there is disagreement about the application or interpretation of an industry level agreement, this is called grievance arbitration. Often these disputes are solved in the industrial tribunal without the arbitrator. It is also possible that parties at the firm level are not members in the industry level organisations and encounter problems in agreeing upon work conditions or pay. Firm and industry level arbitration is called interest arbitration. The arbitrator frequently takes part in these negotiations.

According to Sarkko (1980), the arbitrator has two tasks to fulfil. First, the arbitrator should prevent industrial conflicts by promoting better industrial relations. The second task is to act as a chairperson in the actual negotiations between parties.

\footnote{Normally, only some two-thirds of employees are affected by the central agreement, but the comprehensive incomes settlements in 1995 and 1997 affected almost 100 per cent of the labour force.}
The arbitrator gets involved in the negotiation after a strike or a lock-out is announced. A prerequisite for industrial action is that there is no collective agreement in force\(^9\). The announcement should be given to the office of the arbitrator and to the opposite party fourteen days before the action is planned to take place. The motivation of this requirement is to give enough time for the arbitrator to operate. If the planned action threatens "essential general interests", the Ministry of Labour has a legal right to postpone it for two weeks\(^10\). The decision is made after a recommendation by the arbitrator.

After filing a notice of strike or of any other form of work stoppage, the arbitrator takes whatever action he considers suitable. The parties are obligated to attend negotiations chaired by the arbitrator and they are required to supply any information the arbitrator deems necessary. After receiving "information on all aspects of the dispute", the arbitrator should induce the parties to determine the main matters of the dispute precisely and to restrict them to as few as possible. Then the arbitrator seeks to bring about a compromise, suggesting such concessions and adjustments as appear appropriate and fair. If the parties fail to end up with a compromise, the arbitrator may present a draft settlement to them and recommend to accept it within a short time limit (Kauppinen 1988). According to Sarkko (1980) the arbitrator makes his own proposal in about 80 per cent of cases\(^11\).

One very important feature of the arbitration process should be noted. The parties are obligated to attend negotiations, but they are not obligated to accept the proposal made by the arbitrator. In this sense Finnish wage arbitration differs from the North American counterparts. However, there are two crucial features transferring power to the arbitrator. First, there is only one way to start a legal strike: parties are obligated to announce the industrial action to the arbitrator and to take part in the negotiation chaired by the arbitrator. Secondly, the arbitrator normally does not make a proposal for an agreement until the very last moment before the work stoppage is due. By postponing the offer, the arbitrator is forcing the trade union and the employer to a *take it or leave it* situation. There are always

---

\(^9\)In politically motivated strikes this is not the case: these are normally not announced officially. That is not a problem for the arbitration because in political strikes there is very little to negotiate or agree: the strikes end like they start, without any formal notice.

\(^10\)This does not apply to political strikes or solidarity actions without any selfish purpose.

\(^11\)The statistics of public arbitration are not published anymore. This is perhaps because the number of cases has declined drastically. This is mainly due to fewer industrial disputes.
some fixed costs when the strike or the lock-out starts: it is difficult to stop a strike after it has begun. Thus the parties should think very carefully of accepting the offer by the arbitrator.

The arbitrator has at least some power in the wage setting. The timing of the proposal and the time limit for the answers may make further negotiations before the work stoppage almost impossible. Sometimes, when the solution of the dispute is quite near, the parties can agree on postponing the stoppage for further negotiations. The possibility to continue negotiations after a rejection of the draft settlement is considered undesirable. Both parties may take the proposal by the arbitrator as a minimum for the agreement and that may postpone the compromise. Hence, the arbitrator tries to make his offer as late as possible. The topic is discussed in more detail in Section 4.6.
4.3 Arbitration in North America

In the United States, arbitration is based on law, but it is run as private business. Actually, it is quite a large industry. It is estimated that there were over 100,000 arbitration cases in the United States in the realm of labour-management relations in 1985 (Bloom and Cavanagh 1987). The North American practices have been widely studied and the bulk of theoretical studies are done in that context. Hence, it is natural to contrast the Finnish arbitration system with its North American counterparts.

Arbitration is used in disputes ranging from labour issues to legal questions. Here the main emphasis is on labour market disputes. However, one aspect of legal disputes is worthwhile to point out. An arbitrator is typically selected at least to a degree by mutual agreement of the parties. According to Ashenfelter (1987), the key determinant of the parties’ preferences is usually the extent of the arbitrator’s experience in deciding related cases (see also Bloom and Cavanagh 1986a). In courts, however, juries are selected because they have little experience with the type of dispute at hand. Specifically, they are expected not to have any previous knowledge of the particular case in process. This suggests that arbitration is appealing because a smaller risk is involved in resolving the dispute than would be in a court decision. Later it will be argued, however, that the arbitration is riskier than agreeing.

4.3.1 Forms of arbitration

Most labour contracts\textsuperscript{12} in the United States have an arbitration procedure to settle disputes that arise in the course of administration of the agreement (grievance arbitration). The most common form of interest arbitration in North America is the so-called conventional arbitration. In conventional arbitration, the arbitrator is free to set any wage he or she desires and the parties have to honour the decision.

The use of arbitration is in some cases based on law. This is typical in the case of public sector wage disputes where the right to strike is restricted, e.g.

\textsuperscript{12}In 1980, approximately 97 per cent of collective bargaining agreements in the private sector covering 1,000 or more workers include grievance procedures with arbitration (Farber and Bazerman 1989).
in bargaining between municipalities and fire brigades, police corps etc.\textsuperscript{13}

The use of arbitration can also be based on mutual agreement. For example, Ashenfelter (1987) mentions the practice in New Jersey. The parties opt for conventional arbitration when they agree mutually to adopt this procedure, while final-offer arbitration is used otherwise.

In \textit{final-offer arbitration} the arbitrator’s behaviour is restricted. He or she should choose one of the last offers by the parties if the negotiation breaks down. The idea behind this procedure is to make extreme demands more costly. The parties are forced to take into account the trade-off between the probability of winning and expected size of the victory. After some experience of the use of conventional arbitration in the public sector, it was argued that the arbitrators tend to simply split the difference between the last offers by the parties. If that kind of behaviour is recognised by the parties, they will adopt a strategy to demand as much as possible to maximise their utility. Thus the conventional arbitration procedure does not encourage compromises.

A strike imposes costs on parties directly in form of lost wages and production. Although arbitration may also impose direct costs (Ashenfelter and Currie 1990), the primary channel through which final-offer arbitration imposes costs on parties is fundamentally different. It is the uncertainty concerning the arbitrator’s award combined with the risk-aversion of parties which is hypothesised to make final-offer arbitration a costly alternative. (Farber 1980)

Ashenfelter and Currie (1990) also mention \textit{tri-offer arbitration}, which has been in use in Iowa since 1974. In this form of arbitration, the arbitrator may also choose a fact-finder’s position.

In Canada, the right to strike is limited. Both in the private and in the public sector it is necessary to go through some form of non-binding third-party mediation before a strike or a lock-out can be legally declared (Currie and McConnell 1991). In this sense the system resembles the Finnish one. About one-third of the contracts were negotiated under the compulsory arbitration statutes. Most collective bargaining statutes allow two parties to agree voluntarily to use binding arbitration. Conventional arbitration is practically the only form of arbitration in use.

Very often in economic analyses it is implicitly assumed that the dis-

\textsuperscript{13}According to Freeman (1986), by 1981 twenty States had specified that arbitration is used to determine the terms of labour contract among public sector employees where the parties cannot agree. According to Ashenfelter and Currie (1990), half of States have arbitration statues covering some group of public employees. The final-offer arbitration is in use in six States (Farber 1980).
pute is one-dimensional: there is only one parameter, e.g. the wage rate, in the agreement. There is another possibility leading to one-dimensional bargaining: if utilities of the parties depend on parameters of agreement in exactly the same manner, the problem reduces to one dimension. Both cases are very unlikely. Finnish collective agreements normally consist of roughly one hundred pages of text. Moreover, the negotiating parties value working conditions and pay increases very differently. According to Crawford (1981) issue-by-issue final-offer arbitration is widely used in the United States. In this case, the arbitrator is permitted to design his or her settlement from the components of the parties’ final offers. As an alternative solution to the problem, Donn (1977) and Crawford (1979) suggest a procedure called multiple final-offer arbitration\textsuperscript{14}. In this form of arbitration, the parties are asked to submit more than one offer. If the parties agree, the agreement becomes final. Otherwise, the arbitrator announces which agent has made the most suitable offer. Then the other agent must choose one of the first agent’s offers. As result, the solution is more likely to be at the Pareto-frontier.

4.3.2 Conventional vs. final-offer arbitration

The discussion of advantages of different arbitration schemes begun when final-offer arbitration was started in the 1960’s. The mechanism for settling disputes determines not only the terms of agreement in all cases but also the probability of reaching an agreement without using the dispute settlement mechanism (Farber and Bazerman 1986). Arbitration procedures can be ranked by three criteria: the quality of the settlements they generate when negotiations break down, their freedom of bias and the extent to which they generate negotiated settlements (Crawford 1981).

Crawford (1979) showed that in a zero-sum setting the most commonly used obligatory arbitration schemes are equal if the award by the arbitrator is certain and known to the parties. In conventional arbitration, whatever negotiated settlement was proposed by one party, the other party would always do better to wait for the arbitration award. Hence, all outcomes would coincide with the arbitrator’s will. Assuming that the arbitrator selects the final-offer closest to his or her notion of a fair settlement, the same result holds also for final-offer arbitration. Whatever offer is made by one party, the other could always do better to propose something between the offer and the arbitrator’s most preferred wage. Hence, both parties are

\textsuperscript{14}A similar procedure is used in Eugene, Oregon (Crawford 1979 and 1981).
led to agree on the arbitrator's preferred wage or see it imposed by the arbitrator. (Ashenfelter and Bloom, 1984)

An outside observer would interpret this as a success of arbitration. The parties are eager to agree and costly strikes can be avoided. However, this is illusory because the very nature of bargaining is different. The parties are forced to follow the line of the arbitrator and they agree only to avoid direct costs of the arbitration.

If preferences of the arbitrator are stochastic or not perfectly observed by the negotiating parties, there is a distinction between the arbitration systems. Also, if the arbitrator extracts information on the case from the offers of the parties, the system of arbitration matters.

As Currie points out, it is difficult to draw any inference about the effect of arbitration on wages simply by comparing wages under different forms of arbitration. States with a high wage level may be more likely to adopt arbitration statutes (Currie 1989, see also Currie and McConnell 1991). The same criticism applies to the argument put forward by Bloom and Cavanagh (1987). They point out that arbitration usage rates are not notably different in those States with conventional arbitration provisions to those with final-arbitration provisions. There have been several attempts to avoid the aforementioned sample selection bias and the question is studied by laboratory and field studies. According to Farber and Bazerman (1989), final-offer arbitration leads to higher negotiated settlements rates.

As already noted, final-offer arbitration was developed in response to criticism of conventional arbitration. It was argued that conventional arbitration awards tend to be compromises between the parties' final positions. If the arbitrators really split the difference between the last offers, the resulting situation would be very unfavourable (Feigenbaum 1975 and Feuille 1975). Parties would have an incentive to maintain polar positions in the negotiation to influence the arbitration award most favourably.

Bloom (1987) lists several reasons why the arbitrator would split the difference between offers. First, making a mechanical compromise between the parties may be an optimal strategy for the arbitrator who wants to show fairness. Second, it is a potential way to shirk. Third, if the final offers contain useful information on the case, it is wise to use that information. Nonetheless, it is unlikely that an arbitrator can extract any information from the offers without any reference to exogenous data. If this is the case, the optimal arbitration awards will not be simple compromises, but rather functions of both the offers and the facts.

Bloom also mentions a strong opposite argument. If the parties can forecast the offer by the arbitrator, their negotiations will take place over a
settlement that is near the expected offer by the arbitrator. For example, if the arbitrator is favourable to the trade union, the last offers are centred around the expected arbitration award. The arbitration decisions may appear to be mechanical compromises of the parties’ final positions, although the causal relationship may be reverse.

If the behaviour of the negotiators is really affected more by the arbitrators’ expected award than the award is affected by the offers of the negotiators, there are also other important consequences. First, if final-offer arbitration is more effective in solving the disputes, this is only because the arbitrator is more effective in dictating his or her preferences to the negotiating parties than in other forms of arbitration. Second, as Farber (1980) points out, there is a trade-off between the incentive for negotiations and the quality of arbitrated awards. Maintaining a high degree of uncertainty to promote reaching negotiated settlements is a good mechanism for reducing reliance on final-offer arbitration. However, if the agreement is not reached, uncertainty will result in extreme awards.

The problem at hand is difficult to study empirically. The relationship between the last offers and arbitration awards may be simultaneous for aforementioned reasons. Bloom (1987) and Farber and Bazerman (1986) used simulated cases to avoid simultaneity bias and asked real arbitrators to offer awards. Bloom concluded in his statistical study that the arbitrators do give weight to the facts, but different arbitrators act so differently that the weight tends to show up as random noise.

According to Farber and Bazerman (1986), conventional arbitration awards seem to be weighted averages of the facts and the offers. The weights seem to be dependent on the quality of the offers. Because final offer awards are affected by facts and offers very similarly than the awards in conventional arbitration, Farber and Bazerman conclude to the arbitrators determine their awards without reference to the particular form of arbitration in use.

4.3.3 Empirical analyses

In addition to the theoretical studies, there are several empirical studies available. Chelius and Dworking (1980) studied salary disputes in major league baseball in the United States. They conclude that final-offer arbitration is an effective strike substitute and impasse solution technique.

---

15See also Crawford (1982a).
16However, it can be noted that extreme offers do not exist. This is evidence against "splitting-the-difference" behaviour.
4.3. Arbitration in North America

Burgess and Marburger (1993) concluded, however, that most arbitrated settlements in baseball differ significantly from settlements negotiated by players of comparable value. This evidence suggests that bargaining parties retain substantial freedom to negotiate their own settlements. On the other hand, this means that most final-offer awards are of "low quality" in the sense that they lie outside the range of negotiated settlements\(^ {17}\).

Currie and McConnell (1990) conclude in their study of public sector contracts in the United States that if the aim of collective bargaining legislation is to reduce disputes, granting workers the right to strike is the most effective policy. While the incidence of strikes is the same under arbitration and the right to strike, arbitration hearings are more common under arbitration systems. Strikes are the shortest under compulsory arbitration and longest when strikes are legal. They note also that the form of collective bargaining legislation seems to have no effect on the negotiated wage increase. Hence, it is possible that substituting compulsory arbitration for the right-to-strike would be welfare-improving\(^ {18}\).

Currie and McConnell (1991) got similar results using Canadian panel data of public sector contracts. They found out, however, that wages are higher under compulsory arbitration than under a right to strike or without any collective bargaining legislation.

Currie (1991) studied contracts between teachers and school boards in the Canadian province of British Columbia. She found out that the variance of arbitral wage rates is systematically lower than the variance of wage rates which were freely negotiated. Currie's interpretation of her finding is that arbitrators co-ordinate their awards.

It is known that the arbitration awards are not evenly distributed in the binding arbitration. Ashenfelter (1987) and Ashenfelter and Bloom (1984) presented data from New Jersey police disputes. According to the data, employers have won only about one-third of the final-offer decisions during 1978-1980. According to Ashenfelter, this is not difficult to understand. Using the mean award from conventional arbitration cases as a benchmark, he argues that the offers of workers have been closer to the mean of arbitrators' preferences. Ashenfelter and Bloom (1984) noted that final-offer arbitration leads to lower average wage increases than does arbitration un-

\(^ {17}\)It is interesting to note that this observation does not raise any criticism against arbitration by the authors. The same comment applies to Faurot's and McAllister's study (1992).

\(^ {18}\)Currie and McConnell point out, however, that they have looked at only one dimension of the cost of disputes: the length of strikes. Moreover, they have neglected all other aspects of contracts than the wage change.
der conventional rules. They concluded that what the union bargainers
gave up, therefore, by way of a decrease in the mean award under final-offer
arbitration, they may have made up by a reduction in its variability. Brams
and Merrill (1991) suggested that one way to account for this conservative
behaviour by labour is to postulate that labour, but not management, per-
ceived itself as receiving a bonus for winning. If winning per se is attractive
enough, the union would behave like observed.

In the public sector of Iowa\textsuperscript{19}, the employer side has won in two-thirds
of the cases. However, also in Iowa the winning party has been closer to
the recommended awards elsewhere. According to Ashenfelter, these results
support the arbitrator exchangeability hypothesis\textsuperscript{20}.

Bloom and Cavanagh (1986a) showed some empirical results on choosing
arbitrators. Trade unions and employers seem to have similar preferences.
They are in favour of lawyers, more experienced arbitrators, and arbitrators
who seem to have favoured their side in the past. However, the parties do
not agree on every issue. The employers' side has a strong preference for
having economists as arbitrators while the trade unions are against them.

\textsuperscript{19}The data covers the years 1976-1983.

\textsuperscript{20}The arbitrator exchangeability hypothesis states that arbitrator decisions contain
an unpredictable component. The unpredictability is the key element of the \textit{ex ante}
acceptability of awards. Ashenfelter argues that the arbitrators with predictable bias
are ruled out by the selection process. Only "random" enough arbitrators are accepted
by the parties.
4.4 Towards understanding arbitration

4.4.1 What is the right wage level? - A micro-economic view

The topics dealt with in arbitration vary from simple wage questions to complex issues in working conditions, in working time and in workers’ rights to participate in management. The negotiation issues may be multidimensional even in pure wage setting; there may be different wages for different groups of workers or perhaps a special wage for weekends and night shifts. And also, when piece rates are in use, rules for calculating the rates are needed. These rules can be quite knotty.

In what follows, however, the analysis is restricted to one-dimensional wage questions. The choice is justified because wage questions are difficult enough to understand and analyse even in the one dimensional case. The interplay between wage and working conditions and other issues can be analysed successfully only after the more simple questions are answered.

In this study I attempt to argue that it is fruitful to analyse different wage setting institutions by asking which purpose a certain institution serves. In the context of the Finnish wage arbitration system, one may ask if the system serves some general principles of good labour market performance. Answering that question requires discussion on the features of desired performance. In the present section I will ask what is the right wage level? The nature of the problem forces us to restrict our analysis to micro-economic questions. The growth aspects or business cycle considerations are now bypassed and the question is dealt with as a micro-economic bargaining problem.

An intrinsic element in wage determination is that the size of the cake to be divided depends on actions taken by the firm and the workers. If a co-operative solution is not reached, production may prove impossible or output may turn out to be unsatisfactory. The problem is very often analysed with the concepts of the Pareto-frontier or the bargaining set, see Figure 4.1. The axis in the figure are $V$ and $U$, utilities of the firm and the worker. Depending on the case, the utilities can also be defined as aggregates. $V$ may denote utility of the employers’ association and $U$ aggregate utility of the workers or the trade union.

In the most simple case, utilities can be written as a function of wage
level $W^{21}$. We assume that $W$ is selected from the real axis, and that the utility functions are continuous and twice differentiable. If no agreement is reached, no production takes place and the utilities of the parties are determined outside of the model. The firm’s reservation utility is denoted by $V_s$, and the reservation utility of the workers is $U_s^{22}$. The Pareto-frontier is a border for the bargaining set. Points at the Pareto-frontier, i.e. at the contract curve, can be calculated by solving the following problem for every $V$.

$$\max_{W} \quad U(W)$$

s.t. $V(W) \geq V$ \hspace{1cm} (4.1)

---

21In general, the utilities may depend on other factors influenced by wage level, e.g. on employment and on the production level. For example, the utility of the trade union may be $U = U(W, L(W))$, where $L(W)$ is labour demand. Function $U(W)$ is then a reduced form of $U(W, L(W))$.

22Binmore, Rubinstein and Wolinsky (1986) discuss the proper interpretation of reservation utilities. They argue that the utilities should reflect the utility during the dispute (when time preference matters) or the utility if the bargaining breaks down (when risk of negotiations breaking down matters).
By definition the Pareto-frontier is efficient, i.e. it is impossible to increase utility of an agent without decreasing utility of another agent. However, if no agreement is reached, utility allocation \((V_s, U_s)\) would take place inside the Pareto-frontier.

In what follows, it is assumed that the utility functions satisfy the following properties\(^{23}\).

\[
U'(W) > 0, \quad U''(W) \leq 0
\]

\[
V'(W) < 0, \quad V''(W) \leq 0
\]  

(4.2)

Then the Pareto-frontier

\[
\Omega \left( \underline{V} \right) \equiv \arg \max U(W)
\]

s.t. \(V(W) \geq \underline{V}\)

(4.3)

has the following properties

\[
\Omega' \left( \underline{V} \right) < 0, \quad \Omega'' \left( \underline{V} \right) \leq 0.
\]  

(4.4)

As a consequence, the bargaining set is convex.

The nature of the reservation utilities implies that no allocation with \(V\) smaller than \(V_s\) or with \(U\) smaller than \(U_s\) are agreed upon. There are still innumerable alternatives because the utility functions are continuous and the wage is selected from the real axis.

Which point at the contract curve or at the Pareto frontier is reached? The question can be stated positively as well as normatively. How are the points selected in the real cases? Which point should be selected in an ideal case? Bloom’s and Cavanagh’s notion ”... maximise the disputants' welfare ... ” requires defining how the preferences are aggregated (see page 78). What is the right wage?

The questions raised are very complex and have been discussed widely in social philosophy. The solutions for the problem offered by the axiomatic co-operative game theory are the most valuable here.

Nobel laureate John Nash introduced the \textit{axiomatic approach} by specifying a list of properties that a bargaining solution should satisfy (1950). If there is a unique solution that satisfies a given list of axioms, i.e. theoretical requirements, then the solution is said to be \textit{characterised} by the axioms (see introduction in Conley and Wilkie, 1994).

\(^{23}\)If functions \(U(W)\) and \(V(W)\) were linear, the Pareto-frontier would be a straight line.
Nash required that the solution must satisfy four conditions. First, the solution must be _individually rational_. This axiom means that outcomes inferior to the disagreement outcome are ruled out. The second axiom requires that the solution should be invariant to _positive affine transformations_ of the utility functions\(^{24}\). The third axiom requires _symmetry_: the outcome should not favour one of the bargainers. Fourth, the solution should be _independent of irrelevant alternatives_. This implies that if the bargaining set is reduced such that the original solution point still belongs to the set, the original point should remain as a solution\(^{25}\).

Nash showed that if the set of alternatives is convex\(^{26}\) and if the preferences are "well-behaving", there is a unique solution that satisfies the axioms. The solution maximises the product of the bargainers’ utilities when the utilities are measured as deviations from the utilities obtained when the noncooperative outcome is realised. In a context of wage bargaining the problem can be formulated as follows\(^{27}\).

\[
\max_{W} \quad (U(W) - U_{s})(V(W) - V_{s})
\]

The maximand is called Nash product\(^{28}\) and utility pair \((V(W^{*}), U(W^{*}))\), where \(W^{*}\) solves Problem (4.5), is called Nash-point.

The Nash axioms and the corresponding solution is not the only possible way to solve the bargaining problem. Kalai and Smorodinsky (1975) gave an alternative list of plausible axioms and showed that the solution satisfying the list is equivalent to sharing the benefits resulting from co-operation.

The Nash and Kalai-Smorodinsky solutions to the general bargaining problem are valuable as such. They show that a set of intuitively appealing

---

\(^{24}\) The solution should not change when original utility function \(u(x)\) is replaced by function \(\tilde{u}(x) \equiv a + bu(x), b > 0\).

\(^{25}\) For a detailed exposition, see Osborne and Rubinstein (1994, Chapter 15).

\(^{26}\) The problem can be extended to the non-convex sets, see discussion below.

\(^{27}\) An alternative way to solve the problem is to define the iso-loci for the maximand. The equation

\[(U - U_{s})(V - V_{s}) = \Omega\]

defines implicitly the correspondence between \(U\) and \(V\) for every \(\Omega\). Iso-loci are rectangular hyperbolas. The Nash-point can be determined by selecting the highest \(\Omega\) such that the corresponding pair \((V, U)\) belongs to the bargaining set. The corresponding hyperbola is asymptotic to the Pareto-frontier and the Nash-point belongs to both loci, see Figure 4.1 at page 92.

\(^{28}\) Later, the Nash solution has been generalised for variable bargaining powers and more than two bargainers. See references in Osborne and Rubinstein (1994).
axioms leads to a unique and easily calculable co-operative solution. However, without any further justification, the solution have only normative value and nothing guarantees that the solution actually is realised. Co-operative solution concepts would be more appealing for theorists if one found plausible noncooperative grounds for the solutions. However, the ultimate test for relevance of any theoretical concept must be empirical.

If a plausible noncooperative game with an equilibrium corresponding to the co-operative solution exists, it is said to implement\(^{29}\) the solution concept. Nash (1953) himself started to examine noncooperative grounds for his solution concept. He showed that there is a noncooperative game with an equilibrium corresponding to his axiomatic solution. The bargaining game was quite artificial and, worse still, had multiple equilibria. Moreover, Nash's procedure requires that the designer of the noncooperative game knows the utility functions of the agents. This is quite a strong assumption. It is preferable to design the rules of the noncooperative game such that for any profile of the agent’s utilities unknown to the designer, the equilibrium payoffs are unique and coincide with the allocation determined by the axiomatic solution (Conley and Wilkie, 1994).

The Nash Programme\(^ {30}\), i.e. showing that the axiomatic solution to the bargaining problem can be implemented by constructing dynamic noncooperative games, has advanced very remarkably since publication of the seminal paper by Nash\(^ {31}\). The problem of multiplicity of equilibria can be avoided by requiring that the strategies of the players should be sub-game perfect. In short, the agents should have an incentive to follow the equilibrium strategies during the course of the game. If this perfectness is required, there are games with a unique noncooperative equilibrium that corresponds to the co-operative bargaining solution\(^ {32}\).

\(^{29}\)The concept of implementation is discussed in Section 4.5.

\(^{30}\)See, for example, Binmore’s article in Binmore and Dasgupta (1987) or Binmore, Rubinstein and Wolinsky (1986).

\(^{31}\)Rubinstein (1982) and Binmore, Rubinstein and Wolinsky (1986) have constructed games implementing the Nash solution and Moulin (1984) implementing the Kalai-Smorodinsky solution. Also generalised axiomatic solutions for non-convex bargaining problems (allocations are chosen from a non-convex set of feasible allocations) have been found. For implementation of Nash extension, see Kaneko (1980), Herrero (1989) and Conley and Wilkie (1994). Other axiomatic solutions for non-convex problems are studied in Anant, Mukherji and Basu (1990), in Howard (1992) and Conley and Wilkie (1994).

\(^{32}\)Binmore, Rubinstein and Wolinsky (1986) show that if the risk that the negotiations break down goes to zero in a game of alternating offers, the limit of the unique sub-game perfect equilibrium is equivalent with the Nash solution discussed above. Moreover, if
The Nash solution or the corresponding implementation games are not intended to be realistic descriptions of actual decision making in society. However, the analysis of arbitration may benefit much from the advances of the Nash Programme.

Analysis of Binmore, Rubinstein and Wolinsky (1986) clearly show the factors that affect the outcome in a noncooperative bargaining game. The more impatient the player is, the less he gets. Similarly, if there is a risk that the "cake" disappears during a bargaining delay, the more risk-averse party will suffer and get a smaller share of the "cake". Moreover, because the outside option, described by utilities $U$ and $V$ in our example, affect the outcome, the bargaining outcomes favour people with favourable starting points. It can be concluded that wealthy people with large amounts of capital can expect a good outcome from any bargaining. They are capable of waiting and they can tolerate more losses than poor people.

The co-operative Nash solution is actually a powerful concept with very strong implications. If the agents in the real world are concerned only with maximising their own welfare and if they interact according to rules of the game, the Nash bargaining solution emerges without any outside intervention. The resulting utility allocation fulfils reasonable criteria and, what is most important, the allocation is efficient. The solution is intuitively appealing because it takes into account relative power of the bargainers and the noncooperative game implementing solution explains the sources of bargaining power. By doing this, the solution escapes many caveats of social choice theory, where normative theories like the Rawlsian maximin principle lack a flavour of realism.

There is, however, a remarkable paradox involved in the implementation games. Noncooperative games implementing the Nash solution and other co-operative solutions have shown the factors affecting the allocation resulting in bargaining. By doing so, they have also shown why the implementation of the co-operative solutions in the real world is so difficult or even impossible. Implementation games assume complete information, i.e. the bargainers are expected to know the preferences of the opposite party. However, in the real world that kind of an informational requirement is seldom satisfied. Informational asymmetry and the results of implementation games imply that the parties have an incentive to signal that they are patient, risk-neutral and that they have good outside options available. As a consequence, the bargainers have strong incentives not to follow the rules of the implementation games.

---

the delay between the offers goes to zero, a similar result is obtained.
4.4. TOWARDS UNDERSTANDING ARBITRATION

Though micro-economic theory has proposed good candidates for the right wage level, it is likely that they are difficult to realise in the real world. The outcomes may be inefficient \textit{ex post} or there is risk that the bargaining breaks down and a labour market dispute emerges. The latter option is causes \textit{ex ante} inefficiency. Reasons for inefficiencies are discussed more carefully in the next subsection.

4.4.2 Reasons for labour disputes

Divergent expectations or asymmetric information?

Analysis of the arbitrators’ behaviour cannot go very far without asking what is the ultimate reason for labour market disputes (Farber 1980, Farber and Bazerman 1987). More generally, economists ask why are there delays in agreeing and why are sometimes mutually beneficial agreements not reached (Kennan and Wilson 1993).\footnote{The relation between labour and capital cannot fully be understood if it is studied only as a trade between two independent agents. History does matter, the parties are fundamentally dependent on each other, there are intrinsic asymmetries in the relation, and what is the most important, the wage relation is also a power relation. However, the topic of this dissertation is not the long term "class struggle" between labour and capital, but "everyday life" of labour and capital. Hence, the issues raised by aforementioned authors are the relevant ones.}

Without arbitration labour market disputes are solved by industrial action (Farber and Bazerman 1986). The action normally ends with an agreement, but sometimes the cake disappears before it is divided; the firm goes into bankruptcy before the deal is reached. Arbitration is not a definitive solution to the problem; strikes exist though they may be illegal.

Farber and Bazerman (1987) listed five possible reasons for labour disputes. First, the parties may have divergent and relatively optimistic expectations regarding the ultimate outcome if they fail to agree\footnote{An often cited early reference is Hicks (1963).}. If the contract zone, i.e. the range of wages which both parties prefer to disagreement is empty, there is no way of reaching an agreement. The contract zone is determined by expectations of the disagreement outcome and the risks involved in disputing\footnote{If the parties were risk-averse and had identical expectations about uncertain arbitral settlement, then there would be a non-empty contract zone (Ashenfelter and Currie 1990, Farber and Katz 1979).}. Assuming that systematic differences in expectations are independent of the type of arbitration scheme, then the size of the \textit{identical-expectations contract zone} should be a proper measure of
the likelihood of an agreement in actual cases where expectations may differ. However, the calculations of Farber and Bazerman (1989) showed that these zones are larger in conventional arbitration, but the settlement rates are higher under final-offer arbitration. This is evidence against the divergent expectation theory\textsuperscript{36}. On the other hand, Neale and Bazerman (1983) show that in experimental negotiation situations negotiators systematically overestimate the probability that they will be successful in arbitration\textsuperscript{37}.

Divergent expectations are also challenged at the theoretical level. Kennan and Wilson (1993) reviewed recent studies of asymmetric information. According to them, persistently differing expectations can explain any behaviour except the absence of learning. Furthermore, if the parties learn from past mistakes, the divergence disappears. According to the authors, different expectations can be reduced to different sets of information (see also Ashenfelter and Currie 1990). Hence, the reason for a bargaining impasse must be found in the existence of private information.

Kennan’s and Wilson’s main argument is that the need to find out private information of the opposite party is the fundamental reason for bargaining delays or labour market disputes. The uninformed party may “screen” or the informed party may “signal” information. The authors conclude (Kennan and Wilson 1993, page 92):

Situations in which the parties’ expected payoffs exceed the obtainable gain are usually interpreted as indicative of divergent expectations. Here and more generally, however, they reflect merely that the parties condition their expectations on different information.

Learning is the second point on Farber’s and Bazerman’s (1987) list of the potential explanations of disputes. However, they are not convinced about the significance of asymmetric information. They refer to the observation of sizes of identical-expectations contract zones and the assumption that the size of the zone is a measure of cost of disagreeing. The larger size of the zone under conventional arbitration is not consistent with the fact that there is less disagreement in final-offer arbitration. Learning in

\textsuperscript{36}Moreover, Bloom and Cavanagh (1987) noted that the steady-state rate of arbitration usage seems to vary between 15 and 30 per cent in States with compulsory interest arbitration laws. According to the authors, it is unlikely that the bargainers would consistently be overly optimistic about the size of an arbitration award in the context of what is essentially a repeated game.

\textsuperscript{37}The average estimate for the probability to win was 68 per cent. In reality, only 50 per cent of the last offers can win.
the form of disputes is costlier in conventional arbitration and thus disputes should exist less frequently than in final-offer arbitration. As a conclusion, Farber and Bazerman state that asymmetric information is not a sufficient explanation of disagreement.

The aforementioned discrepancy in interpreting importance of asymmetric information has a crucial implication. In the literature of labour economics, all forms of labour market disputes have been seen as inefficiencies. Work stoppage causes losses in production, bad industrial relations reduce effort and care, an inadequate quality of production and delays ruin the market for products etc. Kennan and Wilson differ radically from this traditional labour economics view. According to them, a wasteful conflict is in some situations a necessary implication of each party's incentive to exploit the advantage of its private information. This kind of thinking has its roots in Myerson (1979). He demonstrated that ex post efficiency can be incompatible with the ex ante incentives of the parties in the bargaining.

In an imaginary command economy, all trades known to have positive gains could be dictated and the costs of delays could be avoided. Creating this kind of environment is a controversial and unrealistic project. Kennan and Wilson argue that in the larger social context, inefficiencies caused by delays and failed agreements are costs imposed by the reliance on the privileges of private property when in fact the parties have different information and some monopoly power. For example, strikes are tolerated in some countries to ensure that workers are not disadvantaged.

Kennan and Wilson mention Scandinavian countries explicitly as an opposite example. According to the authors, the larger social interest is invoked to justify occasional intervention in strikes in the Scandinavian countries. Similarly, binding arbitration is dictated by law in most countries for workers in critical occupations and the public sector. (Kennan and Wilson 1993, page 47)

As a conclusion, it can be noted that the argument put forward by Kennan and Wilson is an interesting standpoint, but it cannot be taken literally. In specific, I think Kennan and Wilson have gone too far in their conclusions in three respects.

First, the right to strike and the binding arbitration dictated by law are extreme options, but not the only alternatives. The use of arbitration is

---

38The same contradiction is also evident in the discussion on reasons for strikes. Strikes can be interpreted as results of mistakes or to the contrary, strikes can be seen as information revealing tools. For references and empirical results, see McConnell (1989).
often based on mutual agreement. In many cases, the arbitration laws are at least tacitly accepted by the trade unions.

Second, the observation that in the Scandinavian countries the right to strike is limited does not take all aspects of the situation into account. As Therborn (1992) and Rowthorn (1992) argue, the Scandinavian countries are characterised by corporatist structures\(^{39}\). Incomes policy is an important part of wage determination and the trade unions are accepted as negotiation partners in a wide range of political issues. Hence, limitations in the right to strike are agreed upon at least in an implicit way. Similarly, as I pointed out in Section 4.2, the Finnish arbitration system is accepted by the trade unions.

Third, informational superiority of conflicts, claimed by Kennan and Wilson, can be questioned. In principle, the arbitration systems can be designed to be more efficient in revelation of private information. For example, in the following sections it will be argued that the Finnish arbitration system creates a similar incentive to agree and make concessions that does the threat of work stoppages. Moreover, in the Finnish system it is always possible to obtain the same information that is available if there is no arbitration. Both sides of the conflict may check "the hard facts" of dispute by rejecting the offer by the arbitrator and by initiating a work stoppage.

**Other reasons for strikes**

The third possible explanation for labour market disputes, mentioned by Farber and Bazerman (1987), is commitment\(^{40}\). Bloom and Cavanagh (1987) point out that the political pressure of the members makes concessions costly and commitment by leaders of the trade union credible. Also, it has been argued that bargaining may end with an impasse because an arbitrator is useful as a scapegoat (Farber 1980).

The fourth explanation refers directly to the structures of arbitration systems. The parties may avoid any concessions in the bargaining because they fear that possible concessions are used as evidence in the arbitration hearings. Farber’s and Bazerman’s (1989) finding that offers in conventional arbitration are much further apart than offers in final-offer arbitration is supporting evidence for the argument.

In final-offer arbitration, the effect of the offer on the arbitration award is

\(^{39}\)See also Calmfors and Driffill (1988) or other articles in Pekkarinen, Pohjola and Rowthorn (1992). Alesina and Perotti (1997) and Summers, Gruber and Vergara (1993) are more recent studies.

\(^{40}\)The seminal papers are Schelling (1956) and Crawford (1982c).
more salient to the parties than in conventional arbitration. This "salience theory" is the fifth on Farber’s and Bazerman’s (1987) list. The authors do not believe that the players in conventional arbitration situations can see through the system and structure the game. The predictions of the standard economic model, which assumes that the parties fully understand the game, are likely to be flawed.

I find it difficult to believe that the negotiators are less informed about the nature of strategic interaction than, say, monopolistic firms at the product market. Wage negotiations are complex, but the same people are engaged in them from year to year; many of them do it on professional basis. Why should "bounded rationality" be more important in wage bargaining than elsewhere? If Farber’s and Bazerman’s critique is correct, it applies to a wide range of issues.

There is also some experimental\textsuperscript{41} evidence of the topic (Farber, Neale and Bazerman 1990). They organised 114 undergraduate students to negotiate artificial labour contracts and the economic model predicted the behaviour of the negotiators quite poorly. The parties reached an agreement in some cases where the economic model would have suggested that there would be no contract zone. In addition, when the parties reached agreement where there was a contract zone, the outcome was not always within the zone. The authors conclude that the development of more complete models of negotiation and bargaining that blend economic and behavioural theories is an important area for further research. I think that this conclusion is plausible.

\footnote{The validity of experiments may be challenged. However, Olson, Dell’Omo and Jarley show (1992) that arbitration awards in single-issue cases are substantially the same as awards by the same arbitrators in actual cases. However, the awards differ in multidimensional cases.}
4.5 Arbitration as an implementation issue

In Subsection 4.4.1, I argued that the Nash co-operative solution is a natural candidate for a wage agreement. It takes into account efficiency, one possible definition of fairness and is "realistic" because it takes into account relative powers of the agents involved in the bargaining. Nevertheless, it is quite unlikely that the Nash solution can ever be implemented if the informational assumptions of the implementation games are not realised. The noncooperative models used in implementation of co-operative solutions showed that the agents would have an incentive to signal that they are patient, risk-neutral and that their utility is high during any conflict. Hence, the agents have an incentive not to follow the rules of the game.

The problem of asymmetric information can be seen as an even more fundamental question. In Subsection 4.4.2 the argument put forward by Kennan and Wilson was discussed. They argued that asymmetric information is an intrinsic element in wage setting and that the expected losses in production caused by work stoppages should be tolerated. Delays in bargaining and all other types of labour disputes are useful because they reveal important information. Thus Kennan and Wilson are critical of wage arbitration and any other form of public intervention in wage setting.

The message of Kennan and Wilson can be put as follows. There is a trade-off between \textit{ex ante efficiency} and \textit{informational efficiency}. The authors seem to put much weight on informational efficiency and claim that work stoppages are the only way to screen fundamental information.

In the study at the hand, I would like to take a different route. Taking informational problems seriously, I would like to ask if the Finnish wage arbitration system is useful just because it gives the parties an alternative way to signal their strength. There is no artificial imperative not to start a work stoppage, but many cases are solved without a strike or lock-out. Perhaps the Finnish system can be seen as a compromise between \textit{ex ante} efficiency and informational efficiency?

Before embarking on the analysis of concrete models of the Finnish wage arbitration system, the question of wage setting under incomplete information is analysed from a wider perspective. Wage setting can be seen as a social choice situation.
4.5.1 Mechanism design: basic concepts

Social choice as a function of preferences

A social choice function describes what economic allocation or any other social decision should look like. It is a function because it should, in general, reflect the fundamentals of society. Specifically, preferences of the members of society, at least when they are involved in decision-making, should affect the choice.

Consider a general setting where independent agents should make a collective choice from set $X$ of possible alternatives$^{42}$. There are $I$ agents, indexed by $i = 1, \ldots, I$. The agents differ from each other by their type $\theta_i$, $\theta_i \in \Theta_i$. The type of an agent defines his preferences over alternatives in $X$, i.e. agent $i$'s utility function is $u_i(x, \theta_i)$, where $x \in X$. Note that the sets $\Theta_i$, $i = 1, \ldots, I$ may vary between the agents and thus the setting may well be asymmetric.

In the wage arbitration case $X$ is a set of feasible wage outcomes. There are two agents: the workers and the firm. Alternatively, the arbitrator can be seen as a third player.

In the general social choice situation, the setting is assumed to be of incomplete information$^{43}$. The types of agents are drawn from a commonly known prior distribution $\phi(\theta_1, \ldots, \theta_I)$. Each agent $i$ observes privately parameter $\theta_i$, but the observation is not made prior to the choice. The distribution function $\phi(.)$ and the utility functions $u_i(x, \theta_i)$ are of common knowledge among the agents.

Preferences of the agents depend on the realisations of $\theta = (\theta_1, \ldots, \theta_I)$. Hence, it is natural to require that the social choice should reflect the realisations; otherwise the choice cannot be efficient. Such an ideal choice can be formalised by defining a social choice function.

Definition 4.1 A social choice function is a function $f : \Theta_1 \times \ldots \times \Theta_I \rightarrow X$, i.e. for each possible profile of the agents’s types $(\theta_1, \ldots, \theta_I)$ there is a collective choice $f(\theta_1, \ldots, \theta_I) \in X$.

Different agents may have very different views about how the decisions should be taken. It is natural for all the agents to want to see that the value

$^{42}$In what follows, the notation of Mas-Colell, Whinston and Green (1995, Chapter 23) is followed.

$^{43}$This assumption will be relaxed in Chapter 4.5.3.
of the social choice function is very sensitive to the value of their personal parameter and independent of the values of the other parameters\textsuperscript{44}.

The problems with social choice functions are well known. The Arrow Impossibility theorem showed that if some very natural and appealing principles are honoured, there is no universal way to aggregate private preferences to social preferences (Arrow 1951). Arrow's impossibility theorem does not, however, postulate that there is no way of making decisions at the social level. Instead, other rules than mechanical aggregation of the private preferences are needed when making social decisions\textsuperscript{45}.

One may ask if the problems associated with implementation of the social choice functions can be overcome by avoiding specifying any mechanism, and instead negotiating an outcome once $\theta$ is known (Moore and Repullo, 1988). Moore and Repullo note that circumventing the *ex ante* choice may be *ex post* efficient, but that in general, there may be *ex ante* inefficiencies. The authors seem not to realise all other problems associated with *ex post* negotiations. Decision-making may in some cases be much easier before the state of nature, i.e. individual characters of the agents, is known. It might be helpful if the agents can credibly commit to follow a certain rule when the individual characters are realised\textsuperscript{46}.

Despite of the fact that the exact form of the social choice function is unclear, it is quite natural to require that it is efficient, i.e. that a social choice function selects a Pareto optimal alternative given the agents' utility functions $u_1(\cdot, \theta_1), \ldots, u_I(\cdot, \theta_I)$. Though the arbitrator may ignore the preferences of the parties, there is no reason to assume that he tries to "punish" the parties by wasting some resources when deciding on the allo-

\textsuperscript{44}The question whether an arbitrator seeks realisation of his own preferences or whether he tries to maximise the parties welfare is actually a question about the form of the social choice function. If the arbitrator chooses between the alternatives according to his own preferences, value of the $f(.)$ function is independent of the preferences of the parties.

\textsuperscript{45}It may be worthwhile to note that voting is very seldom an unproblematic solution to the social choice problems. Black (1948) showed in his seminal paper that voting equilibrium can be guaranteed only if the preferences of the voters are single-peaked. Later is has been shown that the voting equilibrium can be obtained in a multidimensional case only under very restrictive assumptions.

\textsuperscript{46}Even if the labour market parties cannot find a common view, the legislator of the society may be able to do this. I argued that the contractual elements in deciding on the arbitration rules are present in the North American and the Finnish systems. However, it is clear that the rules are also influenced by parliaments and other bodies presenting wider social interest than the labour market parties present. The influence of legislature is evident in the Finnish arbitration system, but is not imaginary in the North American arbitration system either.
cation. However, it will turn out that it may be difficult to bring about, i.e. implement, an ex post efficient outcome. Though the social choice function may be \textit{ex ante efficient}, the ultimate realisations may be ineffective.

**Decentralised decision-making**

Even if the social choice function can be agreed upon, it might be difficult to implement it for two reasons. First, the agents may not be willing to accept the allocation that the function determines. When decision-making is decentralised, it may be difficult to reach an ideal outcome. In principle, this kind of a \textit{decentralised decision-making problem} can be solved by the legal institutions of society. However in reality, the use of force is always difficult.

The second problem is a more fundamental one to solve. When the agents have some private information, implementation of one particular social choice function may be difficult or even impossible. Seminal analyses are Hurwicz (1972, 1973). An agent may have an incentive not to reveal his private information or perhaps to give wrong information about his true type. As a consequence, the information revelation problem constraints the set of social choice functions that can be successfully implemented. This \textit{informationally decentralised decision-making problem} has been an important issue in many fields of economics and decision theory recently. In what follows, this problem of incomplete information is analysed more carefully than other elements in implementation.

**General implementation problem**

In general, it is preferable to have rules for a noncooperative game, such that for any profile of the agent’s utilities unknown to the designer, the equilibrium payoff vector is unique and coincides with the allocation determined by the axiomatic solution (Conley and Wilkie 1994). If such a game exists, it is said to implement the solution concept. Designing the structure that implement the social choice function is called a mechanism design.

The mechanism is designed before the types of agents are known. Thus, in deciding the rules for the mechanism, the social planner has to anticipate the behaviour of the agents. The best way to anticipate the choices is to calculate equilibria of the game created by the mechanism. Hence, the social

---

47 See also the references in Hammond (1979a) and (1979b).

48 The constraint is often called an \textit{individual rationality constraint} and the problem an \textit{incentive compatibility problem}. 
planner knows that if he can choose a suitable set of rules for the mechanism, rational agents may act in such a way that the allocation defined by the social choice function realised. Implementation is unique if there is a single equilibrium in the game created by the mechanism.

**Implementation with multiple equilibria**

We noted that in the context of wage bargaining, a meta-contract of the social choice function may emerge even if there are bitter disputes over every particular wage agreement. It is possible for the workers and the employers to agree upon general terms for a suitable division of income that the firm generates. If a general agreement is achieved, then the labour market parties actually form a collective mechanism designer. One may ask if this possibility reduces the importance of possible multiplicity of equilibria. It might be possible to agree upon the equilibrium which is played if the mechanism can be agreed upon at first hand. However, this conclusion is not always justified because there may be a problem of time consistency\(^49\).

The multiple equilibria problem has been discussed actively in abstract game theory. One approach is to assume that there is an individual, perhaps one of the players, called the *focal arbitrator*. He determines the focal equilibrium in a game by publicly suggesting to the players that they should all implement this equilibrium (Myerson 1991). Even though this suggestion has no binding force, if each player believes that every other player will accept the suggestion by the arbitrator, then each player has an incentive to play according to the suggestion.

In many cases assumption of existence of a co-ordinating individual is too strong, but the multiple equilibria problem looks very different if the arbitrator is involved. The arbitrator may act as a co-ordinator of the wage setting game and indicate an equilibrium that is to be played. Hence, a wider class of social choice function can be implemented uniquely than in a pure

\(^{49}\)In specific, if there is a common time preference and the labour force is very specialised, the workers and the management of the firm may prefer the same rate of investment. If the investments are determined according to profits, then there is a possibility of agreement. However, a possibility to behave opportunistically may ruin the possibility of consensus. That may happen if the trade union or the individual workers worry about the possibility of the firm deviating from the agreement and using the saved money for consumption, not for investments. Classical reference is Lancaster (1973), see also discussion in Pohjola (1983), de Zeeuw (1992), Eriksson (1993) and Seierstad (1993). Grout (1984) and van der Ploeg (1987) analyse an opposite case where the union is able to deviate.
bilateral setting.

There is, nevertheless, no reason to exaggerate the difference between wage bargaining with or without an arbitrator. A focal-point effect may emerge without a middleman if the common culture of the players makes one of the alternative equilibria more natural than the others, or if the payoffs in one particular alternative are much higher than in the others, or if some of the alternatives are very unfair for some players. In addition, the properties of the equilibrium strategies as such may give a focal status to one of the strategies.\footnote{For example, if in a multiple equilibria case there is only one equilibrium with pure strategies and the others are based on randomisation, it would be natural to assume that all the players select pure strategies. Myerson (1991, page 113) points out that the cultural norms may also help to select a focal-point. In a context of wage bargaining, it is clear that in societies with good industrial relations it is quite natural to select an equilibrium with no work stoppage as a focal point if such an equilibrium is available. This holds whether or not an arbitrator is involved.}

\subsection*{4.5.2 Incomplete information}

\textbf{Truthful implementation}

Transmitting information is an essential element of any kind of arbitration. It is interesting to note a related theoretical result from abstract implementation theory, namely the \textit{revelation principle} (Dasgupta, Hammond and Maskin, 1979, Green and Laffont, 1977). According to this principle, if implementation is based on a dominant strategy equilibrium\footnote{The most robust equilibrium concept in game theory is the dominant strategy equilibrium and we have good reason to believe that the agents really follow their dominant strategies if such strategies exist. Dominant strategy equilibrium requires that every agent selects a strategy that gives him at least as large a payoff as any of his other possible strategies for every possible strategy that the other agents might play. It should be emphasised that choosing a dominant strategy does not require any knowledge of the probability distribution of the types of the other agents. And also, it is optimal to act according to dominant strategy even the rivals behave irrationally.} or on a Bayesian Nash equilibrium, any social choice function that can be implemented, can be implemented such that the agents reveal their true preferences.

In a \textit{direct revelation mechanism}\footnote{See Myerson (1991), pages 260-261.}, the agents are asked to announce their types and the social choice function is based on those announcements. \textit{Truthful implementation} requires that at least one of the equilibria of the mechanism is such that every agent tells the truth.

A mechanism is said to implement the social choice function \( f(\cdot) \) in
dominant strategies if there exist dominant strategy equilibria of the mechanism, such that one of the equilibria corresponds to social choice. If a social choice function can be implemented with a mechanism which asks the agents to reveal their types and one of the dominant strategy equilibria of the mechanism is truth-telling, the function is \textit{truthfully implementable in dominant strategies}. It turns out that the if the function \( f(.) \) is implementable in dominant strategies at all, it can be implemented by a mechanism, in which telling the truth is an equilibrium\(^{53}\).

A mechanism implements the social choice function \( f(.) \) in the Bayesian Nash equilibrium if there are Bayesian Nash equilibria\(^{54}\) of the mechanism such that one of the equilibria corresponds to social choice. The social choice function is \textit{truthfully implementable in Bayesian Nash equilibria} if telling the truth is a Bayesian Nash equilibrium of the direct revelation mechanism. All social choices implementable in Bayesian Nash equilibrium are truthfully implementable (Dasgupta, Hammond and Maskin 1979).

These results are of interest for the analysis of wage arbitration. As far as social choice, i.e. wage agreement, is implementable with robust strategies, i.e. with dominant strategies or in Bayesian Nash equilibrium strategies, it can be implemented by saying: “You tell me your type, and when you tell your type, I will play your optimal strategy.” Thus, if the arbitrator is powerful enough to implement the desired outcome, he then is capable of forcing the players to reveal their types.

\textbf{Truthfully implementable social choice functions}

According to the revelation principle, it is enough to study whether the social choice function is truthfully implementable. If the answer is negative, we know that the implementation is not possible with robust equilibria\(^{55}\). Hence, we have a necessary condition for implementability. But this revelation principle does not provide a sufficient condition, since the revelation mechanism may have other, unwanted, non-truthful equilibria (Moore and Repullo, 1988).

\(^{53}\)See Dasgupta, Hammond and Maskin (1979).

\(^{54}\)In Bayesian Nash equilibrium an agent is comparing the expected utilities resulting in different strategies. A strategy of an agent is an equilibrium strategy if the agent cannot increase his utility by changing his strategy given the strategies of the other agents. In that sense the concept is quite robust, though it requires more information and deeper belief on rationality than the dominant strategy equilibrium.

\(^{55}\)The dominant strategy equilibrium and the Bayesian Nash equilibrium are called here robust equilibria.
The requirements for a social choice function to be implementable with robust strategies are quite demanding. The preferences of the agents satisfy the \textit{weak preference reversal property}. According to the property, an agents’ preference rankings should (weakly) reverse when his type changes. This property can be stated in general form using lower contour sets (see Mas-Colell, Whinston and Green 1995, page 872). Gibbard (1973) and Satterthwaite (1975) have proved that under some natural assumptions only \textit{dictatorial social choice functions}\textsuperscript{56} can be implemented with a dominant strategy equilibrium. The \textit{Gibbard-Satterthwaite Theorem} is a strong result and together with the weak preference reversal property it convinces us that arbitration cannot be based on a dominant strategy equilibrium.\textsuperscript{57}

It is thus not surprising that the implementation of Nash and Kalai-Smorodinsky bargaining solutions for general bargaining problem require more elaborated equilibrium concepts and more complex games than is the dominant strategy equilibrium in a static setting. Moreover, what is more important, implementation is based on games where the players observe the type of the opposite side.

Example C.1 in Appendix C highlights the abstract conclusions above. In this simple example, a constant wage is the only social choice implementable in dominant strategies.

Social choice functions implementable in a Bayesian Nash equilibrium can be recognised similarly. Because only truthfully implementable social choice functions are implementable in a Bayesian Nash equilibrium, we need only identify those social choice functions that are truthfully implementable.

Bayesian Nash equilibrium is a more demanding equilibrium concept than the dominant strategy equilibrium\textsuperscript{58}. Hence, it is natural that a wider class of social choice functions can be implemented in the Bayesian Nash equilibrium. For example, if the utility functions are of quasi-linear form, and if the types of agents are statistically independent, there is always an

\textsuperscript{56}A social choice function is dictatorial if there is an agent \( i \) such that \( f(\cdot) \) always chooses one of \( i \)'s top-ranked alternatives.

\textsuperscript{57}All results are not, however, impossibility results. If there are only two alternatives to choose between, majority voting is implementable in dominant strategies, see Mas-Colell, Whinston and Green (1995, page 875). Moreover, if the preferences of the agents are of quasi-linear form, decisions on and financing of common projects can be implemented with dominant strategy equilibria. This is the so-called Groves mechanism, see Groves (1973). Nevertheless, majority voting is not a relevant issue for wage bargaining if the internal life of the trade unions is not dealt with. Moreover, wage arbitration is very difficult to see as decision-making upon a common project.

\textsuperscript{58}All dominant strategy equilibria are automatically Bayesian Nash equilibria, but not \textit{vice versa}. 
ex-post efficient social choice function that is implementable in the Bayesian Nash equilibrium\textsuperscript{59}. These conditions again are too restrictive and the result does not change our main conclusion.

Example C.2 in Appendix C show that implementation of wage agreements might be a very difficult task for the arbitrator. Even if the structure of the example is very simple, the resulting conditions for implementability are quite strict. If non-linear utility functions are used, the conditions would be much more complicated.

One additional problem should be noted. Nothing guarantees that the allocations corresponding to implementable choices are attainable if the agents are free to withdraw from co-operation. Myerson and Satterthwaite (1983) have shown that in a bilateral trade setting with risk-neutral agents, there is no social choice function that is ex-post efficient and gives every agent a nonnegative expected gain from participation in trade and is implementable in the Bayesian Nash equilibrium. Also this may be a relevant point for wage arbitration.

### 4.5.3 Complete information

Until now we have considered implementation in an incomplete information environment. However, much of implementation literature studies the case where each agent observes not only his own preference parameter $\theta_i$, but also the preference parameters $\theta_{-i}$ of all other agents. Though the agents observe all parameters, the observations are not verifiable, i.e. no outsider will observe any of $\theta_i$'s. Hence, implementation cannot be done in the court and, i.e. enforceable ex ante agreements are impossible.

Which kind of social choice functions are implementable under complete information? Clearly, an increase in information cannot restrict the set of implementable social choice functions more than the assumption of incomplete information does. All functions that are implementable in dominant strategies equilibrium are implementable in complete information environment (Mas-Colell, Whinston and Green, 1995).

### Implementation in Nash equilibrium

A corresponding equilibrium for a Bayesian Nash equilibrium in the complete information case is the traditional Nash equilibrium. The mechanism is said to implement the social choice function $f(.)$ in the Nash equilibrium

\textsuperscript{59}See d’Aspremont and Gérard-Varet (1979).
if there is a Nash equilibrium corresponding to the selected social choice. The mechanism is said to strongly implement the social choice function \( f(\cdot) \) in the Nash equilibrium if, for each \( \theta \), every Nash equilibrium of the game induced by the mechanism results in outcome \( f(\theta) \).

If the multiplicity of the equilibria is not seen as a problem, then any social choice function can be implemented if there are at least three agents. However, implementation with multiple equilibria is not satisfactory and strong Nash implementation is required. Maskin has shown in an unpublished paper, that only monotonic social choice functions can be implemented strongly in Nash equilibrium. If \( I \geq 3 \), then a sufficient condition for \( f(\cdot) \) to be Nash implementable is that \( f(\cdot) \) satisfies the monotonicity and the no veto power conditions. Monotonicity and no veto power condition together are quite strong requirements.

Example C.3 in Appendix C explains why implementation wage agreements is not easy even there is complete information.

**Implementation with dynamic games**

The set of implementable social choice functions expands remarkably if the game induced by a mechanism consists of several stages and if off-the-equilibrium-path strategies are required to be credible, i.e. if the equilibria

---

60 The following example adopted from Mas-Colell, Whinston and Green (1995) shows this clearly. In the game of announcements, the agents are asked to reveal the true profile \( \theta = (\theta_1, ..., \theta_I) \) to the social planner. It is agreed that if at least \( I - 1 \) agents announce the same profile \( \theta \), outcome \( f(\theta) \) is selected. Assume that the true profile is \( \hat{\theta} \). Structure of the game guarantees that there is no incentive to deviate from announcing true \( \hat{\theta} \), because an individual deviation does not affect the outcome \( f(\hat{\theta}) \). However, announcing any other \( \tilde{\theta} \) is also equilibrium.


62 Moore and Repullo (1990) complete the analysis of Maskin and show that there is a necessary and sufficient condition for an arbitrary social choice rule to be implementable in Nash equilibrium if there are three or more agents. The authors also show that necessary and sufficient conditions for implementability in a Nash equilibrium in the case of two agents are quite demanding.

63 Monotonicity requires that social choice function should reflect the preferences of the agents "continuously" enough.

64 No veto power condition says that if an outcome \( x \) is top-ranked under \( \theta \) by \( I - 1 \) agents, then \( x \) must belong to \( f(\theta) \) (Moore and Repullo, 1990).

65 If \( X \) is finite and contains at least three elements, if all agents have equal set of possible preference rankings and if \( f(\Theta) = X \), then \( f(\Theta) \) is strongly implementable in Nash equilibrium only if it is dictatorial.
rrium strategies are required to be sub-game perfect. Moore and Repullo (1988) show that if there are three or more agents and at least one divisible private good ("money"), then any choice function can be implemented. Moreover, the authors show that if there is a possibility of sufficient free disposal of goods, the conclusion holds even there are only two agents. The authors argue that the conditions are likely hold in "economic" environments, though the classical problems of voting etc. may be inconsistent with the assumptions. Abreu and Sen (1990) have shown that the conditions for implementability with sub-game perfect implementation can be relaxed even more.

The power of multistage games in implementation is well-known in abstract game theory. Axiomatic bargaining solutions can be implemented with dynamic games when perfectness of the equilibria is required. More generally, almost any feasible payoff allocation that gives each player at least his security level can be realised in an equilibrium of the repeated game. These kind of results are known as folk theorems.

The multistage games with (sub-game) perfect solutions are complicated and are quite unrealistic in their structure. This is not a major problem when general conclusions are drawn for an abstract theoretical purpose, but it is difficult to assume that the games are actually used to implement the wage agreements proposed by the arbitrator. Moreover, a sub-game perfect solution requires that the players are capable of seeing the structure of the game and to forecast rationally how the other agents react in each stage of the game.

The most serious problem with implementation in dynamic games is, however, the one mentioned already in Subsection 4.4.1. In a dynamic setting, the agents have an incentive not to follow the rules of the game if there is incomplete information. All agents try to signal strength, and as a consequence, the setting collapses.

In addition to the multistage games with sub-game perfect solutions, other refinements of equilibrium concepts may also solve the problem with multiple equilibria. In general, these solutions require that the players choose their strategies more carefully than in a normal (Bayesian) Nash equilibrium\(^\text{\textsuperscript{66}}\). Though the structures of the games are not as complicated as in the case of multistage games, the refined solution concepts are very restrictive. It can be questioned whether players in the real world are capable

\(^{66}\text{Abreu and Matsushima (1992) show this if there are three or more players using iteratively undominated strategies. If small fines can be levied on the agents, any social choice function can be implemented exactly in the environment with three or more players (Abreu and Matsushima, 1994).}\)
of avoiding iteratively dominated strategies.

4.5.4 Arbitration and implementation

The Nash solution or the corresponding implementation game are not intended to be realistic descriptions of actual decision-making in society. Nevertheless, these studies are of great value as they highlight the factors affecting the division of imaginary cakes. Time and risk preferences of the agents are decisive. It should be remembered that the results of implementation games for the Nash bargaining are obtained in a complete information setting.

If the dynamic games of complicated structure and with detailed rules are allowed, and if the agents observe the same facts, implementation of almost any wage function is possible. Nevertheless, if one looks at more realistic decision-making situations, conditions concerning information and farsightedness of the agents may be too demanding.

In the real world, implementation operates in an incomplete information environment. If the agents are required to use dominating strategies or follow Bayesian Nash strategies, and if the uniqueness of the solutions is imperative, then the set of implementable social choice functions is very small.

Examples given in Appendix C show that it is not enough to ask the labour market parties to reveal their preferences. In a context of wage setting with incomplete information, a constant wage is the only implementable choice if implementation is based on dominant strategy equilibrium. Bayesian Nash equilibrium is capable of supporting a larger set of wage functions if the utility functions of the players are linear in wage.

The results of Rubinstein (1982) and Binmore, Rubinstein and Wolinsky (1986) explain why agents in the real world signal their strength by delaying the arbitration process or by starting a work stoppage. Moreover, it might be very difficult to force the agents to follow the rules of the game. This is a decentralised decision-making problem, referred to already in Subsection 4.5.1, which has very little, perhaps nothing, to do with informational problems. The agents may simply refuse to follow the rules of the game if they think that the rules work against their will.

\[ {\text{67}} \text{Alternatively, the same is true if the multiplicity of equilibria is solved with refinements.} \]

\[ {\text{68}} \text{A similar result is obtained even if there is complete information and the implementation is based on Nash equilibrium.} \]
The discussion of implementation theory gives us an important lesson. If one takes into account the above-mentioned theoretical factors, it is not difficult to understand why implementation of efficient and justified social choices is so difficult in the real world. We are back asking the same question with which the discussion was started. Either unsatisfactory wages are tolerated or a risk of work stoppages are accepted. Third, and the most relevant alternative in the Finnish case, is that wage setting institutions are aimed at producing a compromise between the conflicting targets.
4.6 A model of the Finnish wage arbitration system

The operation of the labour market is affected by the authorities in all industrial countries. To understand why, it is interesting to ask which social or economic function the intervention serves. My hypothesis is that the rules of the Finnish wage arbitration system are designed to reduce the risk of work stoppages. To serve that purpose, the rules of arbitration should imitate the properties, especially the costs, of labour market disputes.

The discussion of the previous section shows that with simple games and simple strategies, complex wage functions cannot be implemented. Hence, it is reasonable to expect that the wage setting game should have a complicated dynamic structure. I would like to suggest that the actual form of the Finnish arbitration system is a compromise between optimal game and the realities of the labour market. Perhaps the peculiar rules of the Finnish arbitration system can be seen as an imperfect version of the even more complex dynamic game capable of implementing efficient wage agreement without any risk of strikes.

In the North American arbitration systems, the decisive power of the arbitrator is guaranteed by the rules of arbitration; the arbitrator sets the wage. The rules of the Finnish system are more complicated. Before entering into the analysis of usefulness of arbitration, I will show by using a simple game theoretical model that also the Finnish arbitrator can influence the wage outcome. Though the model is simple, it captures the main elements of the Finnish arbitration system.\textsuperscript{69}

The negotiation between parties typically starts before the old agreement ends. However, it is quite normal that there is a holdout\textsuperscript{70}, i.e. a period without a collective agreement\textsuperscript{71}. If the disagreement continues and the negotiations do not terminate, one or both negotiating parties announces a strike or a lock-out. The announcement of a work stoppage is done to the

\textsuperscript{69} However, discussion on what might be the objective function of the arbitrator or, more generally, what is the economic function of the whole arbitration procedure, is postponed until Section 4.7.

\textsuperscript{70} See Gu and Kuhn (1998) for thorough discussion about holdouts.

\textsuperscript{71} In the Finnish labour market system there are many levels of agreements. General agreements defining the basic rules of negotiations may still be in force though there is no wage contract in force.
opposite party and to the office of the arbitrator. It should be done at least 14 days before the strike or the lock-out is due. After the announcement, the negotiations continues with the arbitrator, who acts as a chairperson. Arbitration starts as conciliation and the arbitration offer is done as late as possible. This gives more leverage to the arbitrator. The negotiators know that if they do not reach a settlement before and without the arbitration offer, they would be in a take it or leave it situation. A rejection of the arbitration offer leads to work stoppage.

If the work stoppage takes place, there are always some fixed costs. Even if the agreement is reached just after the beginning of the strike or the lock-out, it will take several hours to distribute information to all firms. Firms and industries with continuous production processes are more vulnerable to short strikes than, say, retail businesses or cleaning work. For example, it takes several days to stop a paper mill. When there is threat of a strike, the process of running down the mill is started more than 24 hours before the strike is due. If the strike starts, it is very costly to restore the production process although the strike may last only a few hours. Hence, rejection of the offer by arbitrator is an important move and backing down is difficult. In what follows, it will be demonstrated that the assumption of fixed costs is essential in the model, it helps simplify the model substantially. As a consequence, the model describes more accurately arbitration in the sectors of economy with continuous production processes.

The negotiation period of fourteen days is modelled simply as a pair of simultaneous wage offers (the first phase of the game) and a pair of simultaneous replies (the second phase). If the employer or the trade union accepts the offer by the opposite party, the game ends and a settlement is reached. A special rule is needed for the case when both parties accept the offer by the opposite party. By following the patterns of the North American arbitration studies, it is assumed that wage is then an average of the offers.

If both parties reject the offer by the opposite side, the arbitrator makes his arbitration offer at the very last moment (the third phase of the game). If one or two of parties reject the arbitration offer in the fourth phase of the game, the work stoppage starts, otherwise the arbitration offer becomes a settlement.

\footnote{A large share of all arbitration cases are disputes at industry-level collective bargaining.} \footnote{Note the link with the analysis presented in Chapter 3.}
4.6.1 The economic model and notations

The utility of the trade union is denoted by $U$ and it depends on the wage as follows.

$$\frac{dU(W)}{dW} > 0, \quad \frac{d^2U(W)}{(dW)^2} \leq 0$$  \hspace{1cm} (4.6)

The utility of the employers\textsuperscript{74} is also a function of wage.

$$\frac{dV(W)}{dW} < 0, \quad \frac{d^2V(W)}{(dW)^2} \leq 0$$  \hspace{1cm} (4.7)

In the North American arbitration literature, the preferences of arbitrators are not explicitly dealt with. According to the arbitrator exchangeability assumption, the arbitrator’s awards are designed so that both parties are ready to hire him or her again. In the Finnish system of arbitration, there is no hiring problem as the arbitrator is a civil servant. However, he cannot behave partially. Otherwise the losing party would refuse to participate in arbitration and the whole system would threaten to collapse. As already mentioned, the aim of this model is to show that the arbitrator is capable of affecting the wage outcome. For this purpose, it is enough to assume that the arbitrator’s preferences can be described with the following relation.

$$\left\{ \begin{array}{c}
\text{settlement} \\
\text{without arbitration}
\end{array} \right\} \succ_a \left\{ \begin{array}{c}
\text{settlement in} \\
\text{arbitration}
\end{array} \right\} \succ_a \left\{ \begin{array}{c}
\text{work} \\
\text{stoppage}
\end{array} \right\}$$  \hspace{1cm} (4.8)

To simplify the model, the utilities in the case of a strike are assumed to be fixed\textsuperscript{75}. The trade union’s utility in the case of work stoppage is denoted by $U_s$. For the firm, $V_s$ is used for the same purpose. Equivalent wage levels are defined as follows.

$$u_s = U^{-1}(U_s)$$  \hspace{1cm} (4.9)

\textsuperscript{74}The Finnish arbitration system deals with disputes at the industry and the firm level. Thus we use here the words "employer" and "firm" interchangeably.

\textsuperscript{75}In reality this cannot be the case. Both parties know that their behaviour during negotiation and arbitration affects the beliefs of the opposite party and thus the behaviour of the opposite party during a work stoppage. Also, the outcome of a strike is always uncertain; otherwise it would be very difficult to understand why there are strikes. However, the assumption is justified as a first approximation. The full structural model dealing also with the sub-game starting from the beginning of the work stoppage would be too complex to analyse. In addition, the notion of expected utility in the case of strike is an index describing the expected wage after the work stoppage, utility during the dispute and the cost of uncertainty associated with the duration of the strike and with the settlement after the strike.
\[ v_s = V^{-1}(V_s) \]  

Immediately, it can be concluded that if \( u_s > v_s \), the contract zone is empty and the strike cannot be avoided. In other words, if the strike-equivalent-wage for the trade union is higher than for the firm, there is no room for agreement. Trade union’s offer is denoted by \( w_u \) and the offer by the employer by \( w_f \). The arbitrator’s award is \( W_a \). In what follows, it is assumed that the contract zone exists and that it is larger than a single point, i.e. \( u_s < v_s \).

### 4.6.2 Solution of the game

An extensive form of the game is shown in Figure 4.2. The game starts with simultaneous offers by the parties. It is modelled assuming that the firm moves first and offers \( W_f \). Then\(^{26}\) the trade union demands \( W_u \), without observing the move by the firm. In the second phase the offers become public and there is again a pair of simultaneous moves. The trade union either accepts (denoted by \( A_c \)) or rejects (denoted by \( R_e \)) the offer of the firm. The firm does the same without knowledge of the decision of the trade union. The game ends and the wage is agreed upon if at least one of the parties accept the offer by the opposite side. Otherwise, the arbitrator makes his offer \( W_a \) and the parties respond simultaneously again. The offer \( W_a \) is ratified as a settlement if both parties accept it, otherwise the work stoppage starts. The small circles in the figure are decision nodes and the numbered squares on the right hand side of figure are end nodes. Information sets are depicted with dotted lines.

The game is solved from the end; this guarantees a sub-game perfect solution.

**The fourth phase.** The solution of the sub-game which starts after the arbitrator’s offer is simple if at least one of the parties has accepted the offer by the opposite side. Then strategies of the players have no effect anymore and the wage settlement is \( W_f \) (end nodes 5–8), \( W_u \) (end nodes 9–12) or \( \frac{(W_f + W_u)}{2} \) (end nodes 1–4)\(^{27}\).

If both parties have rejected the offer of the opposite side, the arbitrator’s offer matters. The sub-game ending with nodes 13–16 can be analysed in a normal form, see Table 4.1. If the arbitrator’s offer lies inside the zone, there are two Nash-equilibria in the game. Both either reject the offer or

\(^{26}\) The order of moves can be reversed without any consequences.

\(^{27}\) It was assumed that if both parties accept the offer of the opposite side, the average of the offers is agreed upon.
accept it. However, equilibrium \((Re, Re)\) is not perfect because strategy \(Ac\) dominates for both players. Hence, it can be concluded that the fourth phase of the game ends with the offer by the arbitrator as a settlement if the offer is on the contract zone\(^{78}\).

\textit{The third phase}. As already noted, the arbitrator’s offer does not matter if one or both of parties have accepted the offer by the opposite side. If not, the arbitrator may offer any wage \(W_a\) satisfying

\[ u_a < W_a < v_s, \quad (4.11) \]

and the work stoppage can be avoided. The arbitrator can credibly com-

\(^{78}\)If the offer is at one end of the zone, i.e. if \(W_a = u_a\) or if \(W_a = v_s\), there would be two perfect equilibria. In one of the equilibria, the offer is accepted and in the other, the work stoppage starts. If the offer lies outside the zone, there is only one perfect equilibrium and that is a strike.
**Trade union**

\[
\begin{array}{ccc}
Ac & Re \\
U(W_a) & U_s \\
V(W_a) & V_s \\
\end{array}
\]

**Firm**

\[
\begin{array}{ccc}
Re & U_s \\
V_s \\
\end{array}
\]

Table 4.1: The fourth phase of the game

mit to do this\textsuperscript{79} and in what follows, it is assumed that the \( W_a \) is known beforehand.

The second phase. The negotiators respond simultaneously to wage offers by the opposite side. Again this game can be analysed in the normal form, see Table 4.2. If both negotiators accept the offer of the opposite party, the average realises. If only one of the parties accepts, the offer of the opposite party is agreed upon. The strategy combination \((Re, Re)\) leaves the question in the hands of arbitrator. Because the arbitrator tries to avoid work stoppages and the contract zone is not empty, \( W_a \) becomes a settlement in this case.

The equilibrium of the second phase depends on relative sizes of \( W_u, W_f \) and \( W_a \). Table C.1 in Appendix C (page 169) displays all Nash-equilibria and perfect equilibria of the second phase. As far as \( W_f > W_u \), the size of \( W_a \) does not affect the equilibria, though it affects the best response mappings. In the equilibrium, both parties accept the offer of the opposite side and the average of the first phase offers realises. If \( W_f = W_u \), then the size of \( W_a \) determines which equilibria is perfect. However, all the outcomes are equal in terms of utilities: the firm gets \( V(W_u) = V(W_f) \) and the trade union \( U(W_u) = U(W_f) \). When \( W_f \) is smaller than \( W_u \), the offer by the arbitrator really has an effect. As far as \( W_f \leq W_a \leq W_u \), the only perfect equilibria is \((Re, Re)\). If the arbitrator's offer is weakly superior to the offer of the opposite side for both players, both parties cannot do better than to

\textsuperscript{79}There is no incentive to deviate from the announced \( W_a \).
Trade union

\[
\begin{array}{ccc}
 & Ac & Re \\
Ac & U\left(\frac{W_u + W_f}{2}\right) & U(W_u) \\
 & V\left(\frac{W_u + W_f}{2}\right) & V(W_u) \\
Firm & & \\
Re & U(W_f) & U(W_a) \\
 & V(W_f) & V(W_a) \\
\end{array}
\]

Table 4.2: Second phase of the game

reject the offer of the opposite party. If the arbitrator’s offer lies outside the range of offers by the parties, the party with the offer closer to \(W_a\) is better off.

The first phase. The parties are assumed to decide their offers simultaneously, without knowing the move by the opposite party. Both parties know that their offer affects how the game is played in the latter phases. For example, the trade union knows from Table C.1 that if it demands a lower wage than the employer offers, then in the second phase the average of the offers is agreed upon. Figure 4.3 illustrates the same information that is available in Table C.1. It is easy to see from the table that the best reply mapping for the trade union is

\[
W_u^* \in \{W \in \mathcal{R} \mid W \geq W_a \quad \text{and} \quad W \geq W_f\}. \tag{4.12}
\]

For the firm’s offers which are smaller than or equal to the arbitration award, the trade union may demand any wage higher or equal to the award. For the firm’s offers higher than \(W_a\), the trade union may demand any wage \(W\) such that \(W \geq W_f\). Similarly, the best reply mapping for the firm is

\[
W_f^* \in \{W \in \mathcal{R} \mid W \leq W_a \quad \text{and} \quad W \leq W_f\}. \tag{4.13}
\]

The fixed point of these mappings is the set

\[
\left\{ (W_f^*, W_u^*) \in \mathcal{R}^2 \mid W_f^* \leq W_a \leq W_u^* \right\}. \tag{4.14}
\]
Figure 4.3: The first phase of the game

All points in this set correspond to an equilibrium. Figure 4.4 displays best reply mappings and fixed points of the mappings.

In the equilibrium, the employer side offers a lower or the same wage as the arbitrator is going to offer. Similarly, the trade union demands the same wage as, or higher than, the arbitrator proposes. Only if both demand exactly the same wage, that is only if $W_u = W_f = W_a$, there is a possibility\textsuperscript{80} that the offers are accepted without arbitration. Otherwise there is always an arbitration hearing which ends with the offer by the arbitrator. This is accepted by the parties in the fourth phase of the game.

The model predicts that the threat of arbitration has a strong effect on the parties. The outcome is always such that a strike is avoided and the wage is the same as the arbitrator’s offer, i.e. for the equilibrium wage holds

\[ W^* = W_a. \]  

(4.15)

However, this wage may be agreed upon in various ways and the model does

\textsuperscript{80}There are four perfect equilibria in the second phase. In three of these, the offer of the opposite party is accepted and in one, both parties reject the offer by the opposite side.
not predict how likely these alternatives are. All the equilibria are efficient from the point of view of the labour market parties because no arbitration costs are assumed. If the utility of the arbitrator is taken into account, all the equilibria are not equivalent. According to Equation (4.8), there is a utility loss if the parties cannot agree without the arbitrator’s involvement.

The model analysed in this section is simple and is not intended to be an accurate description of the Finnish arbitration system. The aim of the exercise was to show that also in the Finnish arbitration system, the will of the arbitrator matters if the contract zone is non-empty and the arbitration award is known and there is no uncertainty about it. In this sense, the analysis corresponds to Crawford’s (1979) analysis of the North American arbitration systems.

The main deficiency of the model is that it does not explain why arbitration is useful, i.e. what are the economic or the social functions of arbitration. This question is addressed more carefully in the next subsection.
4.7 The function of the Finnish arbitration system

4.7.1 Background

In Section 4.4, the argument put forward by Kennan and Wilson (1993) was reviewed and commented. The authors argue that the costs of labour disputes must be balanced against their informational benefits. They claimed provocatively that strikes resulting from the right to strike are fair costs of reliance on the privileges of private property. I argued that the authors have gone too far; although strikes transmit relevant information, there is no reason not to allow parties to avoid dispute costs. Voluntary incomes policy is one way of dodging costs associated with work stoppages; in this section it will be argued that the unique features of the Finnish wage arbitration system is the second. Using a modified version of the model presented in Section 4.6, it will be shown that the Finnish arbitration system increases the costs associated with bargaining impasse. However, it helps circumvent the most expensive options, the strike or the lock-out.

The model presented in Section 4.6 was simple. It could explain why the arbitrator may affect wages even though there is a right to strike. However, the fundamental issue of usefulness of the arbitrator still remains. The arbitrator effectively sets the wage and there is no incentive to try to avoid arbitration. Actually, in the negotiations before official arbitration, there was no reason to agree because the arbitration did not cause any costs. In this section this assumption is relaxed and the uncertainty with arbitration is explicitly dealt with.

It is difficult to study consequences of arbitral uncertainty in the model presented in Section 4.6. The problem is technical. Assume that the negotiating parties are risk-averse. Then there are wage levels $u_a$ and $v_a$ defined as follows.

$$u_a = U^{-1}(U_a) \tag{4.16}$$

$$v_a = V^{-1}(V_a). \tag{4.17}$$

$U_a$ and $V_a$ are expected utilities of the parties if the wage is determined by the arbitrator, but stochastically. Immediately it follows from the assump-
tion of risk-aversion that $u_a < v_a$. Because the arbitrator wants to avoid strikes (see preference relation (4.8)), the distribution of arbitration awards is bounded: all awards must satisfy condition (4.11). Hence, it holds that

$$ u_a < u_a < w_a < v_a < v_a, \tag{4.18} $$

where $w_a$ is a mathematical expectation of the arbitration award. Now it can be shown that the best reply mappings are not properly defined. For example, the best reply to $W_j < u_a$ should be the highest wage that is still smaller than $v_a$. However, such a number does not exist.

### 4.7.2 The model and the solution

The game analysed in Section 4.6 can be defined slightly differently such that the above mentioned problem disappears. Figure 4.5 illustrates a modified game. The second phase of the game is deleted and it is simply defined that if the trade union demands the same as or less than the firm offers, the average is agreed upon. Otherwise the case is arbitrated. Also, the uncertainty of the arbitration award is modelled simply by substituting the nature for the arbitrator. This is the so-called Harsanyi transformation (Rasmussen 1989).

From Section 4.6 it is known that the offer by the arbitrator is accepted if it lies inside the contract zone. This is always the case because the distribution of the award was assumed to be bounded so that (4.11) always holds. As a result, the new game with uncertainty can be analysed using Figure 4.6. Points on the 45° line correspond to situations when both parties propose the same wage. Points below the line are equivalent to the situation when the trade union demands a lower wage than the employer offers. Again, the average of the offers is realised. If the offers do not match, and the demand of the trade union is higher than the offer of the firm, the wage dispute is solved in arbitration. As it was shown already, in this case the arbitrator may set the wage.

---

81 Risk-aversion implies that, for the trade union, a deterministic wage $u_a$ is equivalent to a stochastic wage with a higher expected value. Similarly, the firm prefers the deterministic wage to a wage distribution with a lower expected value.
Figure 4.5: Extensive form of the game with uncertainty

It is then easy to see that the best reply mapping for the trade union is

\[ W_u^* \in \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3, \]

where

\[ \mathcal{H}_1 = \{ W \in \mathcal{R} \mid W > W_f \text{ and } W_f < u_a \} \]
\[ \mathcal{H}_2 = \{ W \in \mathcal{R} \mid W \geq W_f \text{ and } W_f = u_a \} \] (4.19)
\[ \mathcal{H}_3 = \{ W \in \mathcal{R} \mid W = W_f \text{ and } W_f > u_a \}. \]

As long as the offer by the firm is lower than the wage equivalent to arbitration award, that is \( u_a \), it is optimal for the trade union to demand a higher wage than the firm offers (set \( \mathcal{H}_1 \)). If the utility of the firm’s offer is equal to the arbitration utility, then the trade union may demand any wage \( W \) such that \( W \geq W_f \) (set \( \mathcal{H}_2 \)). If the offer by the firm is higher than \( u_a \), then the trade union optimally demands exactly the same wage as the firm
Figure 4.6: The outcomes of the game with uncertainty offers (set $\mathcal{H}_3$). Similarly, the best reply mapping for the firm is

$$W^*_f \in \mathcal{K}_1 \cup \mathcal{K}_2 \cup \mathcal{K}_3,$$

where

$$\mathcal{K}_1 = \{ W \in \mathcal{R} \mid W < W_u \text{ and } W_u > v_a \}$$

$$\mathcal{K}_2 = \{ W \in \mathcal{R} \mid W \leq W_u \text{ and } W_u = v_a \}$$

$$\mathcal{K}_3 = \{ W \in \mathcal{R} \mid W = W_u \text{ and } W_u < v_a \}. \quad (4.20)$$

The fixed point of these mappings is the set

$$\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2,$$

where

$$\mathcal{Q}_1 = \{ (W^*_f, W^*_v) \in \mathcal{R}^2 \mid W^*_f \leq u_a \text{ and } W^*_v \geq v_a \}$$

$$\mathcal{Q}_2 = \{ (W^*_f, W^*_v) \in \mathcal{R}^2 \mid u_a \leq W^*_f = W^*_v \leq v_a \}. \quad (4.21)$$

The equilibrium set of the game consists of two subsets. The first possibility is that the trade union demands more than it expects from the arbitration
(\(W_u \geq u_a\)) and the firm offers less or equal than \(v_a\). Then the arbitration starts and the arbitrator sets stochastically \(W_u\) somewhere between \(u_a\) and \(v_a\). The second alternative is that both players demand and offer the same wage as the opposite party. This agreement should be between the wages that the parties evaluate as indifferent to the arbitration.

Both types of agreements are \textit{ex post} efficient. However, arbitrated agreements are inefficient \textit{ex ante}. The parties’ \textit{ex ante} utility as a whole can be increased if wages are set without arbitration. However, there is a co-ordination problem: which wage should be agreed upon. This problem is severe; the firm prefers to set \(W = v_a\) and the trade union prefers the highest possible wage \(W = v_a\). The co-ordination problem requires co-operative behaviour.

### 4.7.3 Solution of the game in a wider perspective

In the Finnish arbitration system, there are features that are designed to facilitate co-operative solutions. The obligatory period of fourteen days for negotiations can be interpreted as an attempt to find a co-operative solution. Also, the arbitrator may act as a co-ordinator before last-minute offers are made. However, as it was pointed out earlier, there is no guarantee: the co-operative solution is sometimes reached before arbitration and sometimes during arbitration, but often the solution is not reached at all.

A mixture of co-operative and noncooperative behaviour cannot be analysed using formal methods of game theory. Nevertheless, the aforementioned situation can be illustrated heuristically, see Figure 4.7. When obligatory arbitration starts, the parties know that if they do not reach an agreement co-operatively, the noncooperative game has the aforementioned co-ordination problem. Hence, it is very likely that it ends with formal arbitration. The risk associated with arbitration means that point \((V_a, U_a)\) is not at the Pareto-frontier. However, it still lies to the right and upward from the original threat point \((V_s, U_s)\). The parties have an incentive to co-operatively choose one point on the Pareto-frontier, i.e. choose wage level \(W^*\) such that \(u_a \leq W^* \leq v_a\). In this kind of heuristic exposition, there is no reason to try to specify exactly how this co-operative outcome is determined. In the figure, an iso-Nash-product curve is drawn; the Nash-solution would be one natural candidate for the solution concept. The Kalai-Smorodinsky solution is an alternative.

The uncertainty related to the arbitration award has an important function. Under certainty, the ”contract zone” before the arbitration hearing consists of a single point, i.e. \(u_a = v_a\). The only possible co-operative so-
4.7. THE FUNCTION OF THE FINNISH ARBITRATION SYSTEM

Figure 4.7: *Ex ante* Pareto-frontier of the wage setting game

...olution would be \( W^* = u_a = v_a = W_a \), where \( W_a \) is the arbitration award. Adding uncertainty changes the situation remarkably. Arbitration with stochastic arbitration awards allows parties to make their deals freely even though work stoppages can be avoided. The only costs are associated with risk-aversion. Strikes and lock-outs would also cause real *ex post* costs.

It should be noted that arbitrator’s choice of type of uncertainty\(^\text{82}\) may affect the wage outcome. By manipulating the probability distribution, the arbitrator may be capable of moving the outside options \( U_a \) and \( V_a \) disproportionately. If the co-operative solution is based on Nash or Kalai-Smorodinsky bargaining, the location of disagreement points affect the outcome.

A recent unpublished manuscript by Manzini and Mariotti (1998) develops a similar argument than the one put forward above. Using the abstract bargaining model of Rubinstein (1982) as a point of departure, the au-

\(^\text{82}\) The results obtained in this section were based on the assumption that the arbitrator may randomise the arbitration award. Equivalently, it is enough that the negotiating parties cannot forecast the arbitration award precisely. Then the firm and the trade union would form an expectation about the arbitration award. Even if the expectations are rational, i.e. even if the expectation errors are distributed around zero, the risk-aversion of the parties causes the existence of a non-empty contract zone.
thors add a possibility of calling an arbitrator in. If both parties of the negotiation agree, the arbitrator is called to divide the cake between the disputants. Though arbitration is never used in a sub-game perfect equilibrium, the possibility of calling the arbitrator in changes the bargaining outcome remarkably. For example, the authors show that if the bargaining costs are zero, any given division of the cake between the players can be implemented. Moreover, if the division is sufficiently balanced (not favouring one of the parties too much), implementation is unique. Though arbitration has a function of increasing the parties’ willingness to reach a negotiated agreement, the analysis of Manzini and Mariotti points out an opposite possibility. In their model, arbitration may generate inefficient delays even when the players have complete information about the behaviour of the arbitrator.
4.8 Concluding remarks

4.8.1 The Finnish system of wage arbitration

In this chapter, I have compared Finnish wage arbitration with its North American counterparts. The similarity of the systems is not evident. The arbitration award in the Finnish system is not binding and there is a right-to-strike, whereas in USA and in Canada the arbitration award is binding and work stoppages are illegal if arbitration is compulsory. However, I have argued that the strategic behaviour of the labour market parties under Finnish arbitration rules based on labour law gives the State Arbitrator a lot of power. In the model analysed, the arbitrator can affect the wage outcome by presenting a last-minute offer just before the work stoppage is due to start. Hence, the arbitrator is capable of dictating any wage level on the contract zone. The result should not be taken too literally, but it gives some support to my hypothesis that the arbitration systems in Finland and in North America have the same features.

The second main theme of the chapter was the analysis of the economic and the social function of the Finnish arbitration system. I started the discussion from an analysis of the Nash bargaining solution. I argued that the games used to implement the solution actually show why the Nash solution is so difficult to reach in the real world. The parties cannot be forced to play according to rules of the implementation games. This fact has very strong implications if there is asymmetric information. The parties would have an incentive to signal their patience and risk-neutrality by delaying their moves or perhaps starting a work stoppage. By doing so, the parties can expect to achieve a bigger share of the utility "cake" to be divided between them.

I argued that the Finnish wage arbitration system can be seen as an alternative way to achieve the same information that is attainable through labour market disputes. However, arbitration is a better way to achieve this because it is helpful in avoiding production losses. The risks which are present in the labour market disputes can be imitated with smaller costs in an arbitration procedure.

The third observation in the chapter was about the structure of the arbitration process. The rules of the Finnish arbitration system, and the games used to analyse it, are complicated. The process is based on proper
timing of the consecutive moves by the parties. The complicated nature of the arbitration processes is easy to understand if one observes the results obtained in implementation theory (Section 4.5). According to these results, only very simple social choice functions can be implemented with static games. If the multi-stage games with sub-game perfect strategies are adopted, also more complex choice functions can be implemented. However, implementability requires that the parties can observe the relevant facts behind the dispute, i.e. that there is complete information. This is very seldom the case.

The comparison between theoretical implementation models and the practice of wage arbitration suggests that the Finnish arbitration system can be seen as an imperfect version of the even more complex dynamic game capable of implementing efficient wage agreement without any risk of strikes. The rules of the arbitration are a compromise between theoretical requirements for a good system and practical possibilities.

The incomplete nature of the system implies that arbitration sometimes leads to a work stoppage. Moreover, it can be questioned whether the actual social choice rules followed in arbitration are efficient. It is also clear that the arbitration award does not always reflect fully the preferences of the labour market parties. However, in general, there are good reasons to believe that the outcomes in the labour market are better with the Finnish arbitration system than without it.

It is very likely that there are less strikes and lock-outs because the labour market parties can compare their "weights" without starting a work stoppage causing production losses. The allocation is not, of course, identical to that which would prevail after a strike, but at the same time the costs associated with a strike are avoided and the parties may also divide this extra utility.

4.8.2 Feedback mechanisms in arbitration

The starting point of my analysis was to show that the arbitrator may affect the wage outcome even in the Finnish system where there is a right-to-strike. The Finnish system was compared with its North American counterparts, where the arbitrator has a right to set the wage. However, it might be useful to note that the power of the arbitrator is limited also in the North American system of binding arbitration. A successful career as a profes-
sional arbitrator requires that the disputants are willing to hire him\textsuperscript{83}. In this sense arbitration and working of the courts differ fundamentally.

The Finnish arbitration system is somewhere between binding arbitration and the court system. The arbitrator is employed by the State and his position is not directly affected by satisfaction of the negotiators. However, it is not as independent as that of the judges. As I have argued, the working of the arbitration system is based on an implicit acceptance by the employers and the trade unions. If the arbitrator loses his reputation of impartiality, the whole wage setting system collapses. Hence, it is clear that the State Arbitrator is replaced if one of the parties demands this to be done. This has never happened and the impartiality of the arbitrator has been questioned only few times\textsuperscript{84}. The right-to-strike feature in the Finnish wage arbitration system is probably one main reason for this. If the arbitration award is favouring one of the disputants, the another may start a work stoppage.

In the binding arbitration system there is a risk of ending up with a lower utility or a lower profit than would prevail in the strike situation. However, even in that system, any arbitration award whatsoever is not possible. The punishments for illegal strikes are limited and the ultimate constraint is the situation at the labour market. If the arbitration award deviates too much from the market wage, production ceases or the workers quit.\textsuperscript{85}

The authority of the arbitrator is limited in all arbitration systems. The form and the strength of feedback may vary, but there are always certain limits to the arbitrator's behaviour. Also in this sense the North American and Finnish systems are not very different from each other.

---

\textsuperscript{83}There are two main selection mechanisms for arbitrators. In both of them, an impartial agency supplies the disputants a list of potential arbitrators. In the Alternate Strike Mechanism, each party alternately cross a name off the list. The last name remaining on the list is appointed as arbitrator. In the Rank-veto Mechanism an impartial agency makes the appointment according to the rankings announced by the parties. In both alternatives the arbitrators have a strong incentive to maintain a reputation of fairness.

\textsuperscript{84}In some disputes concerning female employees there has been criticism, but it has never gone very far.

\textsuperscript{85}There is no information available on the awards in the Finnish system, but it is known that the arbitration awards are not evenly distributed in binding arbitration (see Subsection 4.3.3), but this does not seem to reduce the acceptability of arbitration. Ashenfechter and Bloom (1984) argue that what the union bargainers give up by way of a decrease in the mean award under final-offer arbitration, they may make up by a reduction in its variability. Brams and Merrill (1991) suggest that one way to account for this conservative behaviour by labour is to postulate that labour, but not management, perceives itself as receiving a bonus for winning.
4.8.3 Is binding arbitration an alternative for Finland?

Arbitration procedures have become more common in the United States recently. These are used to determine the terms of labour contracts in almost all branches of the public sector. In the private sector, grievance arbitration is the main application. In some cases, the parties have agreed to solve disputes using an arbitrator.

The attitude towards wage arbitration is mainly positive in North American economic literature; the article by Kennan and Wilson (1993) is a remarkable exception. However, there are only few recommendations to introduce arbitration to a wider class of issues. This can be interpreted to imply that authors are happy with the application domain of arbitration.

From the Finnish point of view, one may ask if we should follow the American way and increase the use of binding arbitration. It is wise to avoid overly strong conclusions, but the following comments are still justified.

First, the state of industrial relations is very different in Finland than in the United States. Even on European standards, union coverage is very high in Finland and has risen during the last years. As I have argued, Finnish wage arbitration complements incomes policy. The arbitration boom in the United States was parallel to the decline of unionisation. Binding arbitration can thus be seen as a substitute for trade union activity. From this point of view, it is quite difficult to see how binding arbitration would fit with other features of the Finnish labour market institutions, which are very different from those in the United States.

Second, the use of binding arbitration schemes in the public sector may be an attractive option, because in many circumstances the employer role of public authority is unclear. In specific, the interest of tax payers is not always well-presented. On the other hand, there haven’t been many bitter, long lasting labour disputes in Finland recently; the present system seems to work. However, one lesson from the American experience can be drawn. If binding arbitration is introduced, it should be based on acceptance of both parties. The awards by the arbitrator are valueless if the trade unions do not accept the system.

Third, the concept of universal use of binding arbitration can be criticised on the following grounds. Many authors have emphasised that arbitrators try to relate their awards to the general wage development; agreements in similar cases act as yardsticks. When the coverage of binding arbitra-
tion approaches 100 per cent, the outside reference level disappears\textsuperscript{86}. It has been argued that the bargainers are likely to take into account the expected award by the arbitrator. If the arbitrators follow the general development and if this is determined by the awards of the other arbitrators, the resulting system would be very strange. It is very likely that there would be multiple equilibria. It is very difficult to see how that kind of system would dominate the existing system.

Fourth, the indeterminacy of universal binding arbitration can be overcome by changing the status of arbitrators. If arbitration is obligatory and the parties cannot choose the arbitrators, the arbitration system may have other yardsticks than "fairness" or "facts" for the basis of the awards. The Australian wage setting system is an example of this kind development\textsuperscript{87}. An important conclusion can, however, be drawn from the Australian experience. If binding arbitration spreads throughout society, it is very difficult to say when arbitration ends and when incomes policy or wage control starts.

\textsuperscript{86}The Economic and Monetary Union of the EU may change the situation. Wage developments in the other EMU countries as well as the inflation target of the ECB may provide a yardstick for the domestic arbitration awards.

\textsuperscript{87}The Australian wage setting system has been based on arbitration tribunals since the beginning of the century (Hancock and Isaac 1992). When a dispute occurs, the arbitration commission is first required to encourage conciliation between the parties. If the parties manage to reach an agreement, the commission will usually register it. But if the parties fail to agree arbitration is compulsory and the award is given by the commission. Either way the outcome is legally binding (Archer 1992). Hancock and Rawson (1993) point out that voluntary agreements were affected by the parties’ awareness of the tribunals’ norms. Although the system was originally very diffuse, there were several local arbitration commissions and it lacked formal co-ordination, it gradually transformed to a system of centralised wage co-ordination. However, the system was reformulated in 1992 and the Australian government seeks a more restricted role for the tribunals.
References


Appendix A

Appendix to Chapter 2

Proof of Proposition 2.1
(a) Comparison of Equations (2.7) and (2.14) shows that if \( M = m \), \( C_{2}^{**} \) is zero and expected profits are equal. From \( C_{2}^{**} (1 - m - C_{2}^{**}) - M + m = 0 \) one may observe that \( dC_{2}^{**}/dM = 1/(1 - m - 2C_{2}^{**}) \). Because

\[
\frac{d}{dM} \left( \frac{1}{2} + \frac{m^2}{2} - M \right) = -1
\]

and

\[
\frac{d}{dC_{2}^{**}} \left( \frac{(1 - m - C_{2}^{**})^2}{2} \right) \frac{dC_{2}^{**}}{dM} = -\frac{1 - m - C_{2}^{**}}{1 - m - 2C_{2}^{**}} < -1,
\]

it is clear that

\[
\frac{1}{2} + \frac{m^2}{2} - M > \frac{(1 - m - C_{2}^{**})^2}{2}
\]

when \( M > m \).

(b) Because an efficient unemployment compensation \( C_{2}^{*} = M - m \) and \( C_{2}^{**} \) is determined according to equation \( C_{2}^{**} (1 - m - C_{2}^{**}) = M - m \), where \( 1 - m - C_{2}^{**} < 1 \), it is clear that \( C_{2}^{**} > C_{2}^{*} \). Then also \( C_{1}^{**} = C_{2}^{**} + m > C_{1}^{*} = C_{2}^{*} + m \). \( \square \)

Proof of Proposition 2.2
(a) The multiplier \( \lambda \) is larger than one because

\[
\lambda = \frac{1 - m - C_{2}}{1 - m - 2C_{2} + \varepsilon} = \frac{1 - m - C_{2}}{(1 - m - C_{2}) + (\varepsilon - C_{2})} \leq 1
\]
can hold only if $\varepsilon \geq C_2$. However, then $\lambda = q\varepsilon/C_2$ is larger than one as $q > 1$. Hence, $\varepsilon^{***} < C_2^{***}$. From Equation (2.26) one may observe that $\varepsilon > 0$ if $C_2 > 0$. Because $M - m > 0$ by Assumption 2.5, $C_2^{***}$ must be positive.

(b) Expected profit is decreasing in $C_2$.

(c) If $\theta \geq m + C_2 - \varepsilon$, $\pi(\theta) = [\theta + \beta(\theta)][\rho(\theta + \beta(\theta)) - m\rho(\theta + \beta(\theta))] - C_2$ is positive only if $\rho(\theta + \beta(\theta)) > 0$. Linearity of the production function implies then $\rho(\theta + \beta(\theta)) = 1$. If $\theta < m + C_2 - \varepsilon$, $\beta(\theta) = 0$ and $\pi(\theta)$ would be negative regardless value of $\rho(\theta)$.

(d) Because the firm is able to set $\varepsilon = 0$, a positive $\varepsilon$ implies that expected profit is higher with than without an insurance policy. \hfill \Box

Proof of Proposition 2.4

(a) Because $C_2^* = M - m$ and $C_2^{***}$ is determined according to

$$C_2^{***}(1 - m - C_2^{***} + \varepsilon^{***}) = M - m,$$

where $1 - m - C_2^{***} + \varepsilon^{***} < 1$, it is clear that $C_2^{***} > C_2^*$. Then also $C_1^{***} = C_2^{***} + m > C_1^* = C_2^* + m$.

(b) Because by Proposition 2.2 $\varepsilon^{***} < C_2^{***}$, $1 - m - C_2^{***} + \varepsilon^{***} < 1 - m$.

(c) Comparison of profit equations

$$\frac{1}{2} + \frac{m^2}{2} - M \quad \text{and} \quad \frac{(1 - m - C_2^{***})^2}{2} - \frac{q(\varepsilon^{***})^2}{2}$$

reveals that if $M = m$, $C_2^{***} = \varepsilon^{***} = 0$ then the profit equations have the
same value. Because \( d \left( \frac{1/2 + m^2/2 - M}{dM} \right) = -1 \) and

\[
\frac{d}{dM} \left[ \frac{(1 - m - C_2)^2}{2} - \frac{q \varepsilon^2}{2} \right]
= \frac{d}{dC_2} \left[ \frac{(1 - m - C_2)^2}{2} - \frac{1}{q} \left( \frac{\lambda C_2}{2} \right)^2 \right] \frac{dC_2}{dM}
= - \left[ 1 - m - C_2 + \frac{1}{q} \lambda^2 C_2 \right] \frac{dC_2}{dM}
= - \left[ 1 - m - C_2 + \frac{1}{q} \frac{\lambda^2 C_2}{1 - m - 2C_2 + \varepsilon} \right] < -1,
\]

\( 1/2 + m^2/2 - M \) is larger than

\[
\frac{(1 - m - C_{2**})^2}{2} - \frac{q \varepsilon^{**2}}{2}
\]

when \( M > m \). \( \Box \)
Appendix B

Appendix to Chapter 3

Symmetric use of workers

The derivation of iso-profit lines was based on the assumption that all workers are used equally in production, i.e. that \( W_1 = \ldots = W_N = W \) and \( E_1 = \ldots = E_N = E \). As a consequence, the model could be solved using aggregate variables. However, the validity of this symmetry assumption was not checked.

Let us define that two workers 1 and 2 are equal if their effort is determined by the same effort function \( E(W) \), i.e. if \( E_i = E(W_i) \) for \( i = 1, 2 \).

**Assumption B.1** The effort function is strictly concave, i.e. \( E'(W) > 0 \) and \( E''(W) < 0 \).

The following proposition can be stated.

**Proposition B.1** Let Assumption B.1 hold and consider a profit maximising firm employing \( N \) workers with equal effort functions \( E(.) \). In the optimum, all workers have identical wages and take the same effort.

**Proof** In the optimum, the wage bill is minimised subject to given production \( Y \). The following Lagrangian function can be formed:

\[ \Lambda = \sum_{i=1}^{N} W_i - \tau \left[ F\left(E_i^{\frac{1}{N}}N\right) - Y \right] \]

\[ = \sum_{i=1}^{N} W_i - \tau \left[ F\left([E_1^{\frac{1}{N}}E_2^{\frac{1}{N}} \ldots E_N^{\frac{1}{N}}]^{\frac{1}{N}}N\right) - Y \right], \quad \text{(B.1)} \]
where \( \tau \) is the shadow price of the constraint and \( E_i = E(W_i) \). From the first order conditions

\[
\frac{\partial \Lambda}{\partial W_i} = 1 - \tau F'(t) \frac{N}{N} E_i^{\frac{1}{N}-1} \frac{dE}{dE_i} \frac{dE(W_i)}{dW_i} \]

\[
= 1 - \tau F'(t) \frac{N}{N} E_i^{\frac{1}{N}-1} \frac{1}{N} E_i^{-1} E_i^\frac{1}{N} \frac{dE(W_i)}{dW_i} \]

\[
= 1 - \tau F'(t) \frac{1}{N} E_i^{\frac{1}{N}-1} \frac{dE(W_i)}{dW_i} = 0, \quad (i = 1, \ldots, N)
\]

can be solved such that

\[
\frac{1}{E(W_1)} \frac{dE(W_1)}{dW_1} = \cdots = \frac{1}{E(W_N)} \frac{dE(W_N)}{dW_N}. \quad (B.3)
\]

Because of concavity of \( E(.) \),

\[
\frac{d}{dW_i} \left[ \frac{E'(W_i)}{E(W_i)} \right] = \frac{E''(W_i)E(W_i) - [E'(W_i)]^2}{[E(W_i)]^2} < 0 \quad (B.4)
\]

and Equation (B.3) can hold only if \( W_1 = \ldots = W_N \). In can be concluded also that \( E_1 = \ldots = E_N \). \( \Box \)

The analysis can be concluded by noting that the use of aggregate variables in the analysis is legitimate if all the workers in the firm are identical in respect of the effort functions. The closer inspection of the proof of Proposition B.1 gives an additional result.

**Assumption B.2** The effort functions are of the form \( E_i = E(W_i, \Omega_i) \) such that

\[
\frac{\partial E(W_i, \Omega_i)}{\partial W_i} > 0 \quad \text{and} \quad \frac{\partial^2 E(W_i, \Omega_i)}{\partial W_i \partial W_i} < 0.
\]

\( \Omega_i \) is a parameter and may reflect any property of an individual worker affecting the supply of effort.

**Assumption B.3** The labour force of the firm consists of different subgroups \( j, j = 1, \ldots, J \). All workers \( i \) (\( i = 1, \ldots, I_j \)) in sub-group \( j \) are equal such that \( \Omega_{j1} = \ldots = \Omega_{jI_j} = \Omega_j \), i.e. all workers belonging to sub-group \( j \) are equal in terms of effort functions.
Proposition B.2 Let Assumptions B.2 and B.3 hold. Then $W_{j1} = \ldots = W_{jJ} = W_j$ and $E_{j1} = \ldots = E_{jJ} = E_j$ for all $j = 1, \ldots, J$.

Proof From Equation (B.3) it is known that effort-wage elasticities are equal for all workers in the optimum. Partial concavity of the $E(\cdot, \cdot)$-function in respect to $W_i$ for all $i$ guarantees that

$$\frac{1}{E(W_i, \Omega)} \frac{dE(W_i, \Omega)}{dW_i} = \frac{1}{E(W_j, \Omega)} \frac{dE(W_j, \Omega)}{dW_j}$$

can hold only if $W_i = W_j$. □

According to Proposition B.2, if the firm is employs several types of workers, the workers belonging to the same group are paid equally and they work similarly, i.e. with the same effort and care. For example, male workers in the firm have a different role in production than female workers if they have different effort functions than female workers.

General utility function

The utility function used in this chapter is a simplified version of the more complex alternative

$$U = \log(WL) - V(\Phi E) - Q(L),$$

where $Q(L)$ measures disutility of hours worked. If $Q'(\cdot) > 0$ and $Q''(\cdot) > 0$, optimal labour supply $L^*$ is a solution to

$$\frac{1}{Q'(L^*)} = L^*.$$

As a consequence, labour supply, i.e. supply of hours, is constant. Hence, it can be neglected in the analysis.
Concavity and other properties of $g(.,.)$ function

Let us define the function $g(.,.)$.

$$g(w - \overline{U}, \phi) \equiv g(w - \overline{U}) - \phi \quad \text{ (B.5)}$$

Concavity of this function can be shown by calculating

$$\frac{\partial g(w - \overline{U}, \phi)}{\partial w} = \frac{1}{V^{-1}(w - \overline{U})} \frac{\partial V^{-1}(w - \overline{U})}{\partial V} > 0 \quad \text{ (B.6)}$$

and

$$\frac{\partial^2 g(.,.)}{(\partial w)^2} = - \left( \frac{1}{V^{-1}()} \frac{\partial V^{-1}()}{\partial V} \right)^2 + \frac{1}{V^{-1}()} \frac{\partial^2 V^{-1}()}{(\partial V)^2} < 0. \quad \text{ (B.7)}$$

The following derivatives are also obtained.

$$\frac{\partial g(w - \overline{U}, \phi)}{\partial \phi} = -1 \quad \text{ (B.8)}$$

$$\frac{\partial g(w - \overline{U}, \phi)}{\partial \overline{U}} = - \frac{\partial g(w - \overline{U}, \phi)}{\partial w} < 0 \quad \text{ (B.9)}$$

$$\frac{\partial^2 g(w - \overline{U}, \phi)}{\partial \phi \partial w} = 0 \quad \text{ (B.10)}$$

$$\frac{\partial^2 g(w - \overline{U}, \phi)}{\partial \overline{U} \partial w} = \frac{- \partial^2 g(w - \overline{U}, \phi)}{(\partial w)^2} > 0 \quad \text{ (B.11)}$$

Aggregation of the effort functions

Logarithmic effort of a firm having $N$ workers is

$$e \equiv \log(E) \equiv \log(E_1^\frac{1}{N} E_2^\frac{1}{N} \ldots E_N^\frac{1}{N})$$

$$= \frac{1}{N} g(w_1 - \overline{U}_1) - \frac{1}{N} \phi_1 + \ldots + \frac{1}{N} g(w_N - \overline{U}_N) - \frac{1}{N} \phi_N. \quad \text{ (B.12)}$$

If $\phi_1 = \ldots = \phi_N = \phi$ and $\overline{U}_1 = \ldots = \overline{U}_N = \overline{U}$ and thus $w_1 = \ldots = w_N = w$ (see Proposition B.1 on page 155), then it is easily shown that

$$e = g(w - \overline{U}) - \phi. \quad \text{ (B.13)}$$
Proof of Proposition 3.2

Assume that the opposite is true, i.e., that there are women working in the firm with \( \lambda_i \) and men working in the firm with \( \lambda_j \) such that \( \lambda_i < \lambda_j \). In Section 3.2 it was shown that profit is a strictly increasing function of the variable \( v = e - \lambda w \), see Equations (3.13) and (3.23). This allows us to write the equilibrium conditions:

\[
\begin{align*}
\epsilon_f - \lambda_i w_f & \geq \epsilon_m - \lambda_i w_m \\
\epsilon_m - \lambda_j w_m & \geq \epsilon_f - \lambda_j w_f.
\end{align*}
\] (B.14)

In the equilibrium, profit of the firm using women should be at least as high as profit of the same firm if men are employed. Hence, the variable

\[v = \epsilon_f - \lambda_i w_f\]

should have at least as high a value as the variable

\[\epsilon_m - \lambda_i w_m.\]

A similar condition holds for firm \( j \) using male labour. From Equation (B.14) one may solve

\[(\lambda_i - \lambda_j)(w_m - w_f) \geq 0.\] (B.15)

If \( \lambda_i < \lambda_j \), then \( w_f \geq w_m \) and thus \( \epsilon_f \geq \epsilon_m \). Note that Equation (B.15) implies \( \epsilon_f - \epsilon_m \geq \lambda_i (w_f - w_m) \), i.e., female workers work more effectively than male workers and are better paid than men. In what follows, we show that neither (a) \( w_m < w_f \) nor (b) \( w_m = w_f \) can exist in the equilibrium. Hence, it is impossible for women to work in a firm with \( \lambda_i < \lambda_j \).

(a) From Proposition 3.1 it is known that \( \tilde{U} > U \). From Equation (B.11) one may note that

\[g'(w - \tilde{U}) < g'(w - \tilde{U}), \quad \forall \ w \] (B.16)

if \( \tilde{U} > U \). Because the firm \( i \) employs female workers with effort \( \epsilon_f \) and wage \( w_f \), the following must hold:

\[g(w_f - \tilde{U}) - \phi \geq g(w_f - \tilde{U}).\] (B.17)

From Equations (B.16) and (B.17) one can easily note that

\[g(w - \tilde{U}) - \phi > g(w - \tilde{U}), \quad \forall \ w < w_f.\] (B.18)

Clearly, \( w_m < w_f \) cannot be an equilibrium. For the firm \( j \) employing male workers, it would be more profitable to hire female workers, see Equation
(B.18). With wage $w_m$, female workers work more effectively than male workers.

(b) From Equation (B.14) it can be seen that $w_m = w_j$ implies that $e_m = e_j$. Consider firm $j$, i.e. the firm employing male workers. Profit of the firm remains constant if effort is reduced by $\varepsilon$ and wage by $\frac{\varepsilon}{\lambda_j}$.

$$e_m - \lambda_j w_m = (e_m - \varepsilon) - \lambda_j (w_m - \frac{\varepsilon}{\lambda_j}).$$

(B.19)

The slope of the female iso-utility line at point $(w_j, e_j)$ is $\lambda_j$ ($\lambda_i < \lambda_j$). Hence, all women working in the firm $i$ are ready to move to firm $j$ if effort is reduced by $\varepsilon$ and wage by $\frac{\varepsilon}{\lambda_j}$. At the same time, profit of the firm $j$ is unchanged. Hence, $w_m = w_j$ cannot exist in the market equilibrium. □

**Equilibrium with two genders, graphical analysis**

The analysis of market equilibrium can be done graphically, see Figure B.1. Let us start the analysis by assuming that utility is $\bar{U}$ for all workers. If this is the case, firms may hire male workers and obtain $(e, w)$-combinations described by function $g(w - \bar{U})$. Function $g(w - \bar{U}) - \phi$ graphs the points attainable by employing female workers. It is easy to see that these two iso-utility curves cannot form the market equilibrium. The male workers’ effort function is on the right and above the female one. All firms would prefer to hire male workers because profit increases towards left and up in the figure. One may conclude that the utilities of different types of workers cannot be equal in the market equilibrium. Moreover, utility of the male worker must be higher than utility of the female worker. Otherwise the male effort function is above the female one and excess demand of male workers prevails.

Because female workers are homogeneous, they must all have the same level of utility in the market equilibrium; this is denoted by $\bar{U}$. Hence $g(w - \bar{U}) - \phi$ is the equilibrium female effort function. Because male workers work more effectively than female workers, it is intuitively clear that firms value them higher. Utility of the male worker is $\bar{U}$, $\bar{U} > \bar{U}$. Locus $g(w - \bar{U})$ shifts down when $U$ is increased because the derivative $\frac{dg(w - \bar{U})}{d\bar{U}}$ is negative, see Equation (B.9) on page 158. The shape of the locus also changes because $\frac{d^2g(w - \bar{U})}{d\bar{U}^2}$ is positive, see Equation (B.11). It can be concluded that the equilibrium locus $g(w - \bar{U})$ is steeper than the hypothetical $g(w - \bar{U})$. Moreover, it
is located below the hypothetical $g(w - \bar{U})$. In Figure B.1 the graphs of $g(w - \bar{U}) - \phi$ and $g(w - \bar{U})$ cross. This is an essential element of market equilibrium. If the graph of $g(w - \bar{U}) - \phi$ is never above $g(w - \bar{U})$, then female workers would not be hired by firms. Similarly, in the equilibrium $g(w - \bar{U})$ should be above $g(w - \bar{U}) - \phi$ at least somewhere.

The next step is to determine $\bar{U}$, the equilibrium male utility. Its intuitively clear that firms that value effort highly, and firms with low $\lambda$ and thus high $\frac{1}{\lambda}$ employ males. At the other end of the firm distribution, workers are female, thus $\bar{U}$ is determined at the middle of the firm distribution. The last firm (counting from the firm having the lowest $\lambda$) still using male workers should be indifferent between using male or female workers. As a consequence, the iso-profit line of the firm at middle should be tangential to locuses of the functions $g(w - \bar{U}) - \phi$ and $g(w - \bar{U})$. If $\lambda_s$ is the efficiency parameter and thus the slope of the iso-profit line of the firm in the middle, then the location of $g(w - \bar{U})$ is determined such that $\lambda_s = g'(w_f - \bar{U}) = g'(w_m - \bar{U})$, where $w_f$ is the wage paid if the workers are female and $w_m$ is the wage paid to male workers. The location of $g(w - \bar{U})$ determines $\bar{U}$ uniquely.

The firms having $\lambda < \lambda_s$ find their optimal $(w, e)$-combinations from the
$g(w_m - \bar{U})$ function, at a higher wage than $w_f$. Similarly, firms having $\lambda > \lambda_n$ select the optimal points from the graph of $g(w_m - U) - \phi$. Firms with flatter iso-profit lines hire male workers and pay a better wage than other firms because effort is important for them. Firms with steeper iso-profit lines hire female workers because effort is not as important for them. Also, it should be noted that women are ready to work at lower utility than men. As a result the highest attainable effort-wage combination for a female worker is $e_f, w_f$ and the lowest effort-wage combination for a male worker is $e_m, w_m$. Our assumptions determine that $e_f < e_m$ and $w_f < w_m$.

**Proof of Proposition 3.7**

(a) There are modern families only if $\mathcal{U} + \bar{U} \leq 2U_n$ holds when there is zero measure of modern families, i.e. only if

$$2\mathcal{U} + \frac{\phi_f}{s} \leq 2\mathcal{U} + 2\frac{\phi_f - \phi_n}{s},$$

see (iii) of Equation (3.61) and Propositions 3.3 and 3.4. The parameter $s$ is determined according to the labour market equilibrium condition. Straightforward calculation using Equations (3.59) and (3.60) renders $\beta \geq \alpha$.

(b) The firms with $\lambda \in [s_1, s_2]$ employ members of the modern families only if

$$2\mathcal{U} + \frac{\phi_f - \phi_n}{s_2} + \frac{\phi_n}{s_1} \leq 2\mathcal{U} + 2\frac{\phi_f - \phi_n}{s_2}.$$ 

According to straightforward calculation, inequality may hold only if

$$\frac{s_2 + s_1}{s_1} \leq \frac{\alpha + \beta}{\alpha}.$$ 

If $\beta$ approaches $\alpha$, inequality holds only if $s_1$ approaches $s_2$.

(c) The share of modern families approaches one only if $\bar{U} + \bar{U} \leq 2U_n$ holds when the share of modern families approaches unity, i.e. when

$$2\mathcal{U} + \frac{\phi_f - \phi_n}{\lambda} + \frac{\phi_n}{\lambda} \leq 2\mathcal{U} + 2\frac{\phi_f - \phi_n}{\lambda}.$$ 

Note that gender gaps are determined according to the end-points of the $\lambda$ distribution. Straightforward calculation gives

$$\frac{\lambda + \frac{\lambda}{\lambda}}{\lambda} \leq \frac{\alpha + \beta}{\alpha}.$$
Note that \( \frac{\bar{X} + \lambda}{\lambda} > 2 \) by definition as \( \bar{X} \geq \lambda \) (see Assumption 3.3).

(d) The results obtained do not depend on government effort \( \phi_g \) nor the parameter \( \gamma \). \( \square \)
Appendix C

Appendix to Chapter 4

Example C.1

Let $f(\cdot, \cdot)$ be a social choice function in a situation with two agents ($I = 2$). Assume that both agents have two possible types: $\theta_1$ is $\theta_1'$ or $\theta_1''$ and $\theta_2$ is $\theta_2'$ or $\theta_2''$. Truth telling is the dominant strategy equilibrium if

$$u_1 \left( f \left( \theta_1', \theta_2 \right), \theta_1' \right) \geq u_1 \left( f \left( \theta_1'', \theta_2 \right), \theta_1' \right) \quad \text{for} \quad \theta_2 = \theta_2', \theta_2'', \quad (C.1)$$

and if

$$u_1 \left( f \left( \theta_1'', \theta_2 \right), \theta_1'' \right) \geq u_1 \left( f \left( \theta_1', \theta_2 \right), \theta_1'' \right) \quad \text{for} \quad \theta_2 = \theta_2', \theta_2'', \quad (C.2)$$

and if

$$u_2 \left( f \left( \theta_1, \theta_2' \right), \theta_2' \right) \geq u_2 \left( f \left( \theta_1, \theta_2'' \right), \theta_2' \right) \quad \text{for} \quad \theta_1 = \theta_1', \theta_1'' \quad (C.3)$$

and if

$$u_2 \left( f \left( \theta_1, \theta_2'' \right), \theta_2'' \right) \geq u_2 \left( f \left( \theta_1, \theta_2'' \right), \theta_2'' \right) \quad \text{for} \quad \theta_1 = \theta_1', \theta_1'' \quad (C.4)$$

where $u_i(x, \theta_i) = u_i' (\theta_i, \theta_{-i})$, $\theta_i$ is utility function of agent $i$. Inequalities (C.1) and (C.2) require that Agent 1 should find telling the truth profitable in all circumstances and Inequalities (C.3) and (C.4) set the same requirement for agent 2.

Inspection of Inequalities (C.1) - (C.4) shows that these are unlikely to hold in the context of wage bargaining. The only exception is the case where
the wage does not depend on types of the agents. Assume that Agent 1 is a worker. If his utility is strictly increasing in wage as it is very often assumed, Inequalities (C.1) and (C.2) can hold simultaneously only if \( f \left( \theta'_1, \theta'_2 \right) = f \left( \theta'_1, \theta'_2 \right) \) for \( \theta_2 = \theta'_2, \theta''_2 \). If the utility of Agent 2, i.e. the employer, is strictly decreasing in wage, it can be concluded that the only social choice that can be truthfully implemented with dominant strategies is a constant wage. Because only truthfully implementable social choice functions are implementable in a dominant strategy equilibrium, a constant wage is the only social choice implementable in dominant strategies.

Example C.2

Let \( f (\cdot, \cdot) \) be the social choice function in a situation with two agents \((I = 2)\). Agent 1 is again the worker and Agent 2 is the employer and \( x \) denotes wage level. Both agents have two possible types: \( \theta_1 = \theta'_1 \) or \( \theta''_1 \) and \( \theta_2 = \theta'_2 \) or \( \theta''_2 \). \( \theta_1 = \theta'_1 \) with probability \( p \) and \( \theta_1 = \theta''_1 \) with probability \( 1 - p \). Similarly \( \theta_2 \) is \( \theta'_2 \) with probability \( q \) and \( \theta_2 = \theta''_2 \) with probability \( 1 - q \). Assume that the utility functions are of linear form: \( u_1 (x, \theta_1) = \theta_1 x \) and \( u_2 (x, \theta_2) = \alpha - \theta_2 x \), where \( \alpha \) is an uninteresting positive constant. Social choice function \( f (\cdot, \cdot) \) is truthfully implementable in a Bayesian Nash equilibrium if inequalities

\[
q \theta'_1 f \left( \theta'_1, \theta'_2 \right) + (1 - q) \theta'_1 f \left( \theta'_1, \theta''_2 \right) \geq \theta''_1 f \left( \theta''_1, \theta''_2 \right)
\]

(C.5)

and

\[
q \theta''_1 f \left( \theta''_1, \theta'_2 \right) + (1 - q) \theta''_1 f \left( \theta''_1, \theta''_2 \right) \geq \theta''_1 f \left( \theta''_1, \theta''_2 \right)
\]

(C.6)

hold for worker (Agent 1) and if the similar inequalities hold for employer (Agent 2).

Existence of truthful a Bayesian Nash equilibrium requires thus that

\[
f \left( \theta'_1, \theta'_2 \right) = f \left( \theta''_1, \theta'_2 \right) = f \left( \theta'_1, \theta''_2 \right) = f \left( \theta''_1, \theta''_2 \right)
\]

(C.7)
or that

\[
\frac{f \left( \theta_1', \theta_2' \right) - f \left( \theta_1'', \theta_2'' \right)}{f \left( \theta_1'', \theta_2'' \right) - f \left( \theta_1', \theta_2' \right)} = \frac{1 - q}{q}
\]

\[
\frac{f \left( \theta_1', \theta_2'' \right) - f \left( \theta_1'', \theta_2' \right)}{f \left( \theta_1'', \theta_2' \right) - f \left( \theta_1', \theta_2'' \right)} = \frac{1 - p}{p}.
\]  

(C.8)

It can be concluded that the existence of a truthful Bayesian Nash equilibrium and thus the implementability of \( f \left( \theta_1'', \theta_2' \right) \) can be guaranteed even if the wage is not constant. However, the probabilities \( p \) and \( q \) restrict the set of implementable social choice functions remarkably. For example, if \( p = q = \frac{1}{2} \), then \( f \left( \theta_1', \theta_2' \right) = f \left( \theta_1'', \theta_2'' \right) \), must hold for all \( \theta_1', \theta_1'', \theta_2' \) and \( \theta_2'' \).

**Example C.3**

Let \( f \left( \ldots \right) \) be a social choice function in a situation with two agents \((I = 2)\). Agent 1 is again the worker and Agent 2 the employer and \( x \) denotes the wage level. Both agents have two possible types: \( \theta_1 \) is \( \theta_1' \) or \( \theta_1'' \) and \( \theta_2 \) is \( \theta_2' \) or \( \theta_2'' \). \( \theta_1 \) is \( \theta_1' \). Types are observable, but not verifiable. Assume that the utility functions are strictly increasing (Agent 1) and strictly decreasing (Agent 2) in wage \( x = f \left( \ldots \right) \). The social choice function \( f \left( \ldots \right) \) is implementable in Nash equilibrium if the inequalities

\[
\begin{align*}
\text{if } x = f \left( \ldots \right) \text{ is increasing in } \theta_1, \quad & u_1 \left( f \left( \theta_1', \theta_2' \right), \theta_1' \right) \geq u_1 \left( f \left( \theta_1'', \theta_2' \right), \theta_1' \right) \\
& u_2 \left( f \left( \theta_1', \theta_2'' \right), \theta_2' \right) \geq u_2 \left( f \left( \theta_1'', \theta_2'' \right), \theta_2' \right) \\
\text{and } x = f \left( \ldots \right) \text{ is decreasing in } \theta_2, \quad & u_1 \left( f \left( \theta_1', \theta_2'' \right), \theta_1' \right) \geq u_1 \left( f \left( \theta_1'', \theta_2'' \right), \theta_1' \right) \\
& u_2 \left( f \left( \theta_1', \theta_2' \right), \theta_2'' \right) \geq u_2 \left( f \left( \theta_1'', \theta_2' \right), \theta_2'' \right) \\
\text{and } x = f \left( \ldots \right) \text{ is increasing in } \theta_1, \quad & u_1 \left( f \left( \theta_1', \theta_2'' \right), \theta_1' \right) \geq u_1 \left( f \left( \theta_1''', \theta_2'' \right), \theta_1' \right) \\
& u_2 \left( f \left( \theta_1', \theta_2''' \right), \theta_2'' \right) \geq u_2 \left( f \left( \theta_1''', \theta_2''' \right), \theta_2'' \right) \\
\text{and } x = f \left( \ldots \right) \text{ is decreasing in } \theta_2, \quad & u_1 \left( f \left( \theta_1'', \theta_2''' \right), \theta_1'' \right) \geq u_1 \left( f \left( \theta_1''', \theta_2''' \right), \theta_1'' \right) \\
& u_2 \left( f \left( \theta_1'', \theta_2'' \right), \theta_2''' \right) \geq u_2 \left( f \left( \theta_1''', \theta_2'' \right), \theta_2''' \right)
\end{align*}
\]  

(C.9) - (C.12)
hold. Strictness of the utility functions dictates that

\[ f(\theta_1', \theta_2') \geq f(\theta_1'', \theta_2') \quad \text{and} \quad f(\theta_1'', \theta_2') \geq f(\theta_1', \theta_2') \]
\[ f(\theta_1', \theta_2') \leq f(\theta_1', \theta_2'') \quad \text{and} \quad f(\theta_1', \theta_2'') \leq f(\theta_1', \theta_2') \]

and

\[ f(\theta_1', \theta_2'') \geq f(\theta_1'', \theta_2'') \quad \text{and} \quad f(\theta_1'', \theta_2'') \geq f(\theta_1', \theta_2'') \]
\[ f(\theta_1', \theta_2'') \leq f(\theta_1', \theta_2'') \quad \text{and} \quad f(\theta_1', \theta_2'') \leq f(\theta_1', \theta_2'') \]

Hence, for the Nash implementable social choice functions hold

\[ f(\theta_1', \theta_2') = f(\theta_1'', \theta_2') = f(\theta_1', \theta_2'') = f(\theta_1'', \theta_2'') = f(\theta_1', \theta_2'). \]  (C.13)
<table>
<thead>
<tr>
<th></th>
<th>Size of parameters</th>
<th>Nash-equilibria</th>
<th>Perfect equilibria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Trade union</td>
<td>Firm union</td>
</tr>
<tr>
<td>A</td>
<td>$W_f &gt; W_u$</td>
<td>Ac</td>
<td>Ac</td>
</tr>
<tr>
<td>B</td>
<td>$W_f = W_u &lt; W_a$</td>
<td>Re</td>
<td>Ac</td>
</tr>
<tr>
<td>C</td>
<td>$W_f = W_u = W_a$</td>
<td>Ac</td>
<td>Ac</td>
</tr>
<tr>
<td>D</td>
<td>$W_a &lt; W_f = W_u$</td>
<td>Ac</td>
<td>Re</td>
</tr>
<tr>
<td>E</td>
<td>$W_f &lt; W_u &lt; W_a$</td>
<td>Re</td>
<td>Ac</td>
</tr>
<tr>
<td>F</td>
<td>$W_f &lt; W_u = W_a$</td>
<td>Re</td>
<td>Re</td>
</tr>
<tr>
<td>G</td>
<td>$W_f &lt; W_a &lt; W_u$</td>
<td>Re</td>
<td>Re</td>
</tr>
<tr>
<td>H</td>
<td>$W_f = W_a &lt; W_u$</td>
<td>Re</td>
<td>Re</td>
</tr>
<tr>
<td>I</td>
<td>$W_a &lt; W_f &lt; W_u$</td>
<td>Ac</td>
<td>Re</td>
</tr>
</tbody>
</table>

Table C.1: The second phase of the game