PUBLIC AND PRIVATE INTERESTS IN REGULATION

Essays in the Law & Economics of Regulation
PUBLIC AND PRIVATE INTERESTS IN REGULATION

Essays in the Law & Economics of Regulation

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van
doctor
aan de Universiteit Utrecht
op gezag van de
Rector Magnificus Prof. Dr. W.H. Gispen
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen
op 5 februari 2003 des namiddags om 16.15 uur

door

Johannes Aleidus den Hertog
geboren op 15 april 1955
te Delft
Promotoren:

Prof. Dr. J.J. Siegers  
(verbonden aan de Faculteit Rechtsgeleerdheid van de Universiteit Utrecht)

Prof. Dr. R.J. Van den Bergh  
(verbonden aan de Faculteit Rechtsgeleerdheid, Erasmus Universiteit Rotterdam)
To Vera, Laurien, and Josefien, - who keep me smiling,

and to Caroline, - who keeps me,
   - who fills me gratitude,
   - who makes me love her.
Preface

This thesis would not have been possible without the help of many others who were so generous with their time, information, and comments. I regret that there is insufficient space to thank all those who richly deserve my gratitude, but I would like to mention here Professor Jacques Siegers and Professor Roger Van den Bergh for their penetrating and valuable criticism, their encouragement and psychological support, to whom I am deeply indebted.

Johan den Hertog
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General Theories of Regulation</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>The Explanation of Deregulation in the Netherlands</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>An Economic Analysis of the (Self-)Regulation of Physicians in the Netherlands</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>Reversed Solidarity in Pension Plans</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>Noncompetition Clauses: Unreasonable or Efficient?</td>
<td>115</td>
</tr>
<tr>
<td>6</td>
<td>Summary and Conclusions</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Summary in Dutch</td>
<td>179</td>
</tr>
</tbody>
</table>
# DETAILED TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
</tbody>
</table>

## Chapter 1 General Theories of Regulation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>1.2</td>
<td>Public interest theories of regulation</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Imperfect competition</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Unbalanced market operation</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Missing markets</td>
</tr>
<tr>
<td>1.2.3.1</td>
<td>Information markets</td>
</tr>
<tr>
<td>1.2.3.2</td>
<td>External effects and public goods</td>
</tr>
<tr>
<td>1.2.4</td>
<td>Undesirable market results</td>
</tr>
<tr>
<td>1.3</td>
<td>Criticism of the public interest theories</td>
</tr>
<tr>
<td>1.4</td>
<td>A more sophisticated version of public interest theories of regulation</td>
</tr>
<tr>
<td>1.5</td>
<td>Private interest theories of regulation</td>
</tr>
<tr>
<td>1.5.1</td>
<td>The Chicago theories of regulation</td>
</tr>
<tr>
<td>1.5.1.1</td>
<td>Stigler</td>
</tr>
<tr>
<td>1.5.1.2</td>
<td>Peltzman and others</td>
</tr>
<tr>
<td>1.5.1.3</td>
<td>Becker</td>
</tr>
<tr>
<td>1.6</td>
<td>Criticism of the private interest theories of regulation</td>
</tr>
<tr>
<td>1.7</td>
<td>Regulation, deregulation and re-regulation</td>
</tr>
<tr>
<td>1.8</td>
<td>Summary</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>The Explanation of Deregulation in the Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Aims and scope of this paper</td>
</tr>
<tr>
<td>2.2</td>
<td>Three theories explaining deregulation</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The public interest theories of regulation</td>
</tr>
<tr>
<td>2.2.2</td>
<td>The Chicago theories of regulation</td>
</tr>
<tr>
<td>2.2.3</td>
<td>The public choice theories of regulation</td>
</tr>
<tr>
<td>2.3</td>
<td>Deregulation in the Netherlands</td>
</tr>
<tr>
<td>2.4</td>
<td>Causes of successes and failures to deregulate</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusions</td>
</tr>
</tbody>
</table>

## Chapter 3 An Economic Analysis of (Self-)Regulation of Physicians in the Netherlands

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>3.2</td>
<td>The (self-)regulation of the medical profession</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Definition of the medical profession</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Regulation and (self-)regulation of the profession of medical</td>
</tr>
<tr>
<td></td>
<td>specialists</td>
</tr>
<tr>
<td>3.2.2.1</td>
<td>Indirect barriers to entry: education and training</td>
</tr>
<tr>
<td>3.2.2.2</td>
<td>Direct barriers to entry</td>
</tr>
<tr>
<td>3.2.2.3</td>
<td>Regulation and (self-)regulation of competition</td>
</tr>
<tr>
<td>3.2.2.4</td>
<td>Reform proposals</td>
</tr>
<tr>
<td>3.3</td>
<td>The economic logic of (self-)regulation of medical specialists</td>
</tr>
<tr>
<td>3.3.1</td>
<td>The logic of regulation: the public interest theory</td>
</tr>
<tr>
<td>3.3.1.1</td>
<td>Provision of accessible health care</td>
</tr>
<tr>
<td>3.3.1.2</td>
<td>Guaranteeing the quality of health care</td>
</tr>
<tr>
<td>3.3.1.3</td>
<td>Macro-economic control of the health care sector</td>
</tr>
<tr>
<td>3.3.1.4</td>
<td>Acceptable incomes and a balanced distribution of income</td>
</tr>
<tr>
<td>3.3.2</td>
<td>The logic of regulation: the private interest theory</td>
</tr>
<tr>
<td>3.3.2.1</td>
<td>Medical specialists as a pressure group</td>
</tr>
<tr>
<td>3.3.2.2</td>
<td>Anomalous features in health care regulations</td>
</tr>
<tr>
<td>3.3.2.3</td>
<td>Competition and market outcomes</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusions</td>
</tr>
</tbody>
</table>

**Appendices**

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regulations and Objectives</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>Statistics on Medical Specialists</td>
<td>82</td>
</tr>
</tbody>
</table>

**Chapter 4**  
Reversed Solidarity in Pension Plans  

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>87</td>
</tr>
<tr>
<td>4.2</td>
<td>The actuarial logic of reversed solidarity in pension plans</td>
<td>88</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Pension benefits in general</td>
<td>88</td>
</tr>
<tr>
<td>4.2.2</td>
<td>The mechanism of reversed solidarity</td>
<td>90</td>
</tr>
<tr>
<td>4.3</td>
<td>The existence and stability of reversed solidarity in pension plans</td>
<td>94</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Informational asymmetry and transaction costs</td>
<td>95</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Imperfect competition</td>
<td>96</td>
</tr>
<tr>
<td>4.3.2.1</td>
<td>Legal and institutional barriers to competition</td>
<td>97</td>
</tr>
<tr>
<td>4.3.2.2</td>
<td>Market failures from adverse selection and opportunistic behaviour</td>
<td>98</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Mutual interest of unions and employers in reversed solidarity</td>
<td>99</td>
</tr>
<tr>
<td>4.3.3.1</td>
<td>Employers and final pay pension plans</td>
<td>100</td>
</tr>
<tr>
<td>4.3.3.2</td>
<td>Unions and final pay pension plans</td>
<td>102</td>
</tr>
<tr>
<td>4.4</td>
<td>Reversed solidarity: a Dutch peculiarity or an international phenomenon?</td>
<td>104</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Outline of the pension schemes in the United Kingdom, Belgium, Germany and France</td>
<td>105</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Preliminary conclusions on the international prevalence of reversed solidarity in pension plans</td>
<td>110</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary and conclusions</td>
<td>112</td>
</tr>
</tbody>
</table>
Chapter 5 Noncompetition Clauses: Unreasonable or Efficient? 115

5.1 Introduction 115
5.2 Traditional legal arguments to restrict noncompetition clauses 117
5.3 Economic explanations of noncompetition clauses 123
5.3.1 Noncompetition clauses to refund training expenditures 124
5.3.2 Noncompetition clauses to protect trade secrets and confidential information 126
5.3.3 Noncompetition clauses and self-selection 128
5.3.4 Noncompetition clauses to prevent opportunistic behaviour and renegotiation 131
5.4 Undue hardship and reputation 135
5.5 Summary and conclusions 137

Chapter 6 Summary and Conclusions 139

6.1 Summary and conclusions of the particular fields of research 139
6.1.1 General Theories of Regulation 139
6.1.2 The Explanation of Deregulation in the Netherlands 142
6.1.3 The Regulation and (Self)-Regulation of the Medical Profession 143
6.1.4 Reversed Solidarity in Pension Plans 145
6.1.5 Noncompetition Clauses: Unreasonable or Efficient? 147
6.2 Overall conclusions on public and private interest theories 149
6.2.1 Comments on public interest theories of regulation 150
6.2.2 Comments on private interest theories of regulation 152
6.2.3 General findings of the research 154
6.3 Policy implications 156
6.4 Future research 159

References 163

Summary in Dutch 179
Introduction

The six chapters in this volume all focus on regulation. Regulation is a broad concept, and for now it may best be defined as the employment of legal instruments by public actors to pursue public and private interests. Regulation may usefully be distinguished in economic regulation and social regulation\(^1\).

Generally speaking, *economic regulation* refers to the regulation of a specific industry and concerns restrictions on price, entry and exit and other dimensions of decision-making by the firm. Regulation of this kind is often justified with arguments based on imperfect competition. Some examples of economic regulation are the regulation of prices or profits in industries with natural monopoly characteristics, the requirement of prior approval to enter or exit a particular market, and the allocation of exclusive broadcasting rights to cable operators. Chapter three in this volume analysis a complex of economic regulations, specifically those pertaining to the market for physician’s services in the Netherlands. This heavily regulated market includes elements such as educational and training requirements, the legal protection of the title and of the profession, price and quality regulations, and regulations on approved forms of organization and cooperation.

*Social regulation* concerns the regulation of common aspects of the behaviour of firms as they may be encountered in a variety of different industries. Social regulation aims to protect the worker, the consumer or the environment and is accordingly subdivided into health, safety and environmental regulation. The typical legal form of regulation is standard setting backed by administrative or criminal sanctions. A common justification of social regulation is the presence of externalities or informational problems. Some examples of social regulations are: the requirement to perform activities without presenting unreasonable risks of injury, the mandatory disclosure of warnings on product labels, and the prescription of the maximum concentration of chemicals in the water supply. In this volume, chapters four and five focus on social regulation. Chapter four examines certain aspects of pension benefits for workers. Chapter five analyses noncompetition clauses, which are part of the labour contract for one in every

five employees in the Netherlands. Chapters one to three focus on general theories of regulation, deregulation, or regulatory developments. Chapter six summarizes and concludes this volume.

Generally speaking, it is possible to differentiate between three branches of research in the law and economics of regulation. First, one branch of economic analysis focuses on the explanation of regulated sectors or industries and the form the associated regulations take. Some typical questions that may be asked are: can we explain regulation of sectors that exhibit natural monopoly characteristics from public interest reasons or from private interest perspectives, who benefits from regulation, why do we find regulation in one competitive sector, but not in another, and why is the use of standards predominant in environmental and health regulations, what explains the use of instruments that are broadly considered to be inefficient? A second line of research concentrates on identifying and measuring the effects of the regulations. For instance, what are the effects of safety regulations, are there fewer accidents or do the regulations have compensatory or adverse effects, what are the effects of environmental regulations on the behaviour of firms or market performance, does the rate of return regulation lower the profits of the regulated industries and what are the effects of regulation on product quality, innovation or economic growth? A third branch of research concentrates on the evaluation of regulation and the design of optimal regulatory institutions and mechanisms. This branch seeks to answer questions such as: is regulation efficient compared with the status quo or with the use of other instruments such as markets or taxes, how should one regulatory regime be transformed into another governance structure or into a competitive regime, what is the comparative efficiency of different legal instruments such as mandatory disclosure, specification standards or prior approval, what type of standard should be chosen and what level of performance

\[\text{\footnotesize \textsuperscript{2} Surveys of the economics of regulation can be found in Aranson (1990), Baron (1995), Joskow and Noll (1981) and Hägg (1997).} \]

\[\text{\footnotesize \textsuperscript{3} For a review, see Ogus (1994a).} \]

\[\text{\footnotesize \textsuperscript{4} Joskow and Rose (1989).} \]

\[\text{\footnotesize \textsuperscript{5} For example Laffont and Tirole, (1993); Gruenspecht and Lave (1989); Baron (1989).} \]
should be demanded, and what are the costs and benefits attached to the use of franchising compared with public enterprises?

In this volume, chapter one presents a review of the literature on the economic theories of regulation. Chapters two, three, and four all focus on the explanation of certain regulations or developments in regulation. Chapter two brings the different theories of regulation to the fore in order to explain why the policy of deregulation in the Netherlands was unsuccessful until the mid 1990s, and why since then the partial or full deregulation of economic sectors has become more common. Chapter three analyses the self-regulation and regulation of the medical profession in the Netherlands in order to determine which of the competing regulatory theories performs best in explaining this complex of regulation. Can we explain price, entry and quantity and quality regulations from public interest motivation, or are private interests the driving forces? Chapter four uses public and private interest theories of regulation to explain reversed solidarity in pension plans. A pension plan is said to exhibit reversed solidarity when people who have no career path contribute to the pension provisions for those who with a career. In general terms, reversed solidarity comes down to blue-collar workers and women paying for the pensions of white-collar workers, most of whom are men. This dimension of pension plans is surprising in view of the fact that their funding is regulated by law and unions occupy half of the seats on pension funds boards.

Chapter five presents an effect-analysis of certain informational problems of the employment relationship and evaluates a number of related solutions. This chapter analysis noncompetition clauses, which are a means of protecting the employer from competition by the employee, for example by prohibiting the employee from working in the same sector or region for a certain period of time after termination of the employment. Courts and legislators often disapprove of such clauses. They are often assumed to be primarily the result of the unequal bargaining position of the employees, and this has prompted proposals for legislation in the Netherlands to severely restrict their use in the near future. However, can it be considered likely that an unequal bargaining position of the employees explains these clauses when they are included in the labour contracts of one in every five employees? What are the effects of the proposed regulation for the employment relationship? The analysis reveals that the regulation of non-
competition clauses is likely to have adverse effects both for the economy and the employees: investments and growth will decline, as will employment and wages.

Throughout this volume, public and private interest theories of regulation will be used as a viewpoint and as an instrument to analyse and evaluate particular regulations or policy developments. Economists once assumed that regulation was a form of government intervention in the market economy in order to pursue the public interest. A fundamental theorem of economics holds that every competitive market equilibrium is efficient. This means that there is no other feasible allocation where all the market participants are at least as well off and at least one is better off, in terms of their preferences. If market imperfections are present and Coasian bargaining and private law remedies offer no efficient solution, the public interest theories of regulation predict that regulation will be instituted when its benefits outweigh its costs. For example, the regulation of the power industry was assumed to result from the monopoly position of power distributors and the need to provide power safely and to provide it to remote and sparsely populated areas. The public interest theories or market failures theories of government intervention were questioned following a number of empirical studies that seemed to refute the original hypotheses of government behaviour. A new complex of theories gradually developed, and they have become known as the private interest theories of regulation, in which regulation results from the interaction of groups pursuing their own interests. For example, the regulation and self-regulation of the professions is hypothesized to result from a trade of transfer of rents to the professions for votes to politicians who are seeking re-election.

There is a debate in the literature about the explanatory power and scientific status of the public and private interest theories of regulation. The final chapter of this volume evaluates some of the aspects of this debate and draws a number

---

7 Boadway and Bruce (1984), chapter three.
of conclusions on the merits and possible applications of the public and private interest theories of regulation. Chapters one to four were previously been published separately in refereed journals or books, and chapter five has been accepted for publication\(^9\).

---

\(^9\) Some of these chapters are reprinted here with minor alterations.
CHAPTER 1  GENERAL THEORIES OF REGULATION

1.1 Introduction

In legal and economic literature, there is no fixed definition of the term 'regulation'. Some researchers devote considerable attention to the various definitions and attempt through systematisation to make the term amenable to further analysis\(^1\). Other researchers, however, entirely abstain from a further definition of regulation\(^2\). In order to delineate the subject and because of the limited space, a further definition of regulation is nevertheless necessary. In this article, regulation will be taken to mean the employment of legal instruments for the implementation of social-economic policy objectives. A characteristic of legal instruments is that individuals or organizations can be compelled by government to comply with prescribed behaviour under penalty of sanctions. Corporations can be forced, for example, to observe certain prices, to supply certain goods, to stay out of certain markets, to apply particular techniques in the production process or to pay the legal minimum wage. Sanctions can include fines, the publicizing of violations, imprisonment, an order to make specific arrangements, an injunction against withholding certain actions, or closing down the business.

A distinction is often made between economic and social regulation\(^3\). Economic regulation consists of two types of regulations: structural regulation and conduct regulation\(^4\). Structural regulation is used for regulating market structure. Examples are restrictions on entry and exit and rules against individuals supplying professional services in the absence of recognized qualifications. Conduct regulation is used for regulating behaviour in the market. Examples are price control, rules against advertising and minimum quality standards. Economic regulation is mainly exercised on natural monopolies and market structures with limited or excessive competition.

---

4. Kay and Vickers (1990); the distinction refers to the Structure-Conduct-Performance approach to the study of Industrial Organization.
Social regulation comprises regulation in the area of the environment, labour conditions (occupational health and safety), consumer protection and labour (equal opportunities etc.). Instruments applied here include regulation dealing with the discharge of environmentally harmful substances, safety regulations in factories and workplaces, the obligation to include information on the packaging of goods or on labels, the prohibition of the supply of certain goods or services unless in the possession of a permit and banning discrimination on race, skin colour, religion, sex, or nationality in the recruitment of personnel.

In the economic theories of regulation, a distinction can be made between positive and normative theories. The positive variant is directed to the economic explanation of regulation and deriving the consequences of regulation. The normative variant investigates which type of regulation is the most efficient. The latter variant is called normative because there is usually an implicit assumption that efficient regulation would also be desirable; for the distinction between positive and normative theories, see the discussion between Blaug (1993) and Hennipman (1992). In the rest of this article, theories will be discussed which are directed to the economic explanation of regulation. These theories can be divided into public interest (nos 2-7) and private interest (nos 8-14) theories of regulation, see for example Ogus (1994). The normative theories will not be discussed further in this overview. Normative theories of regulation generally make a cost-benefit analysis of various regulatory instruments. The following costs may be distinguished:

---

5 For an extensive treatment of the relation between positive economics and policy objectives, see Hutchison (1964).

6 On optimal regulation, see generally Baron (1989), Laffont and Tirole (1993) and Spulber (1989). For a theoretical comparison of public enterprises, regulation and franchising, see Viscusi et al. (1996); for an empirical comparison of public and private enterprises: Board-man and Vining (1989), Borcherdig, Pommerehne and Schneider (1982), Daves and Christensen (1980) and Davies (1971, 1977); but see also Martin and Parker (1997) and Parker (1993) for different results. External effects such as pollution and accidents may be opposed by such instruments as taxes, regulation, liability or negotiations, see Burrows (1999), Shavel (1984a, 1984b), Wittman (1977), White and Wittman (1983) and Weitzman (1974). Information problems may be countered by various legal instruments such as information regulation, prior approval, target standards, specification standards and performance standards, see Ogus (1994b) and Richardson, Ogus and Burrows (1982).
1. the costs of formulating and implementing regulation;
2. the costs of monitoring, enforcement, and maintaining regulation;
3. the costs of compliance with the rules for the industry;
4. the deadweight costs resulting from economic distortions connected with
   nos 1-3 and adverse effects for other policy objectives.

The benefits consist of improvements in the static and dynamic efficiency in the
application of scarce resources. The static efficiency comprises productive and
allocative efficiency. In productive efficiency, production takes place at mini-
mum cost, whereas allocative efficiency implies that resources are put to their
highest valued uses. Dynamic efficiency refers to future improvements in the
application of scarce resources. Through such means as organizational or tech-
nological innovations, fewer resources are necessary in the production of cer-
tain goods. New products and product varieties can also be developed that bet-
ter serve the preferences. Finally, dynamic efficiency refers to the speed at
which markets clear and economies stabilize.

1.2 Public interest theories of regulation

The first group of economic theories of regulation account for regulation from
the point of view of aiming for public interest. This public interest can be fur-
ther described as the best possible allocation of scarce resources for individual
and collective goods. In western economies, the allocation of scarce resources is
to a significant extent coordinated by the market mechanism. In theory, it can
even be demonstrated that, under certain circumstances, the allocation of re-
sources by means of the market mechanism is optimal. Because these condi-
tions are frequently not met in practice, the allocation of resources is not opti-
mal and a demand for methods for improving the allocation arises. One of the
methods of achieving efficiency in the allocation of resources is government
regulation. Government regulation may be efficient when market failures are

7 Arrow (1985).
8 Bator (1958).
9 Shubik (1970); Arrow (1970).
present and private law offers no efficient solution. An efficient allocation of resources will result if the benefits of government intervention outweigh its costs. In the literature, one can find several lists of market failures to justify government intervention\textsuperscript{10}. Among these are: natural monopolies, excessive competition, externalities, informational asymmetries, public goods, trade-cycles, bounded rationality, co-ordination problems, paternalism and redistribution. Summarizing these market failures from the perspective of the operation of a market, government regulation could be interpreted as an efficient instrument to correct imperfect competition, unbalanced market operation, missing markets and undesirable market results. These market failures will now be discussed from the perspective of justification of regulation. The assumption is that the government acts as an omniscient, omnipotent and benevolent regulator\textsuperscript{11}.

\subsection*{1.2.1 Imperfect competition}

In the first place, regulation can improve the allocation by facilitating, maintaining, or imitating market operation. The exchange of goods and production factors in markets assumes the definition, allocation and assertion of individual property rights and freedom to contract\textsuperscript{12}. The guarantee of property rights and any necessary enforcement of contract compliance can be more efficiently organized collectively than individually. Furthermore, the costs of market transactions are reduced by property and contract law.

The freedom to contract can, however, also be used to achieve cooperation between parties opposed to market operation. Agreements between producers give rise to prices deviating from the marginal costs and an inefficient quantity of goods is put on the market. \textit{Antimonopoly legislation} is aimed at maintaining the market operation through monitoring the creation of positions of economic power and by prohibiting competition limiting agreements or punishing the misuse thereof. Imperfect competition can also result from the special characteristics of the production process in relation to the magnitude of the demand in


\textsuperscript{11} Dixit (1996), pp. 4-13.

\textsuperscript{12} Pejovich (1979).
the market. At a given magnitude of demand the cost of production is mini-
mized if the production is concentrated in one company. In that case a natural
monopoly exists. If several companies produce the same quantity of goods, the
unit costs of production would rise. An example of how such a situation arises
is when the production process requires a great deal of capital. In that case, the
fixed costs per unit can continue to decline as production increases. Particularly
in the case of declining marginal costs, average total costs may persistently fall. In such cases it is desirable, from the point of view of productive effi-
ciency to concentrate the production in a single company. A monopolist striv-
ing for maximization of profits will, however, sets prices that deviate from the
marginal costs. The objective of productive efficiency in the production process
then acts to the detriment of the aim for allocative efficiency. Natural monopo-
lies are then either put under control of the state, as has happened in many
European countries, or highly regulated, as has happened for example in the
United States. In the latter case, regulation consists of barring entry to the mar-
ket and the enforcement of profit or price rules that aim to promote allocative
efficiency. In this way, the market results of perfect competition are simulated.
Examples of companies assumed at some time to have possessed the character-
istics of a natural monopoly are railways, electricity distribution, gas and oil
pipelines, telecommunication networks, drinking water distribution and sewer-
age.

1.2.2 Unbalanced market operation

In the second place, regulation may be aiming at the stabilization of markets
and the acceleration of the movement towards market equilibrium. Imbalances
within an economy occur at the level of separate markets and on a macro level.
In separate markets, what is known as destructive or excessive competition can
arise, often as a result of long-term excess capacity. The development of a new
equilibrium can take a long time if the individual participants are in a prisoner’s
dilemma. For all market parties jointly, efficiency is achieved if the existing ex-

---

cess capacity can be rationalized. For an individual producer, however, the 'sunk costs' can mean that it is rational to wait until other suppliers have sorted out the capacity. Because this consideration applies to all producers, the excess capacity can persist for a considerable time. Excess capacity situations can also arise when the production capacity is adjusted to the demand during peak moments or peak periods. Examples are peak loads in the rush-hour (busses, underground railways and trains), during the harvest in agriculture (trucks) and during the tourist high season (touring cars, aircraft). Excessive or ruinous competition can finally also arise in a natural oligopoly. In that case efficiency is achieved if only a few companies supply the market. The small number of companies allows them to react to each other's market strategies, so that among other things, price wars can be waged.

A consequence of excessive competition is not only that the prices sink below the average total costs, but also that the price level fluctuates more widely. This causes insecurity and inefficient decision making on the part of both producers and consumers. Finally, excessive competition can be at the expense of safety and reliability when consumers are not in a position to assess the quality of goods\(^\text{15}\). In the past it was assumed that the situation of excessive competition applied to sectors such as air travel and passenger or goods transport by road or water. For these sectors, business licensing restrictions were devised and the capacity was pruned, sometimes in combination with minimum price regulation. However, modern regulatory theory considers the collection of excessive competition-rationales of government intervention to be 'an empty box'\(^\text{16}\).

Except at the level of the individual sectors, imbalances can also occur on a macro-economic level. Market economies are characterized by a *trade cycle*, the regular alternation of periods of increasing and declining economic activity. In the course of the trade cycle, a self-sustaining process comes about in the goods market that is not compensated by adjustments in the labour market. This arises partly because of lack of information, long-term labour contracts and efficiency wages. Trade cycle policies can be desirable to prevent temporary disturbances to the equilibrium having permanent effects. For example, capital goods of limited usefulness in other market segments can be lost forever in a recession. Fur-

\(^{15}\) Idem.

\(^{16}\) Breyer (1982), pp. 29-32.
thermore, structural unemployment can arise when unemployed workers lose their skill and motivation. Finally, stabilization of the trade cycle can be desirable to prevent the decline of production and employment such that different social groups are unequally affected by the economic rise and fall. Traditionally, trade cycle policies are put into effect together with instruments of budgetary and monetary policy; for an overview of the significance of these instruments and the underlying theories, see Snowdon, Vane and Wynarczyk (1994). Because these instruments are not directed to specific sectors and only take effect after some time, wage and price regulation have been developed in some market economies. To combat a wage-price spiral, governments have for example developed the means to freeze wages and prices for a period of between a half to one year, possibly in designated sectors\textsuperscript{17}.

1.2.3 Missing markets

Public interest theory explains regulation from viewpoints not restricted to imperfect competition and unbalanced market operation. For a number of reasons, markets may not exist for some goods for which the utility or the 'willingness to pay' exceeds the production costs. Markets might not exist as a result of information problems and transaction costs in the case of external effects and public goods. In these cases, regulation can improve the allocative efficiency of the economy.

1.2.3.1 Information problems

In the first place, missing markets can be accounted for by hidden information or an asymmetric distribution of information with respect to prices, quantities or quality of goods\textsuperscript{18}. In this connection, it is useful to make a distinction between 'search goods', for which the quality of a product can be determined prior to purchase, 'experience goods', for which quality only becomes apparent after

\textsuperscript{17} Ogus (1994), p. 300 ff; Breyer (1982), p. 60 ff.

\textsuperscript{18} Hirschleifer and Riley (1979); Stiglitz (1975).
consumption of the good and 'credence goods', for which the quality cannot even be established after consumption\textsuperscript{19}. Examples of each are the purchase of flowers, second hand cars and medical advice, respectively. When it is not possible to establish the quality of goods or services in advance, purchasers will be prepared to pay an average price corresponding with the expectation of quality. Sellers of goods of high quality will not be prepared to offer the goods at that asking price, and will withdraw from the market. The consequence is that the quality of goods traded on the market will decline, as will the price buyers are prepared to pay\textsuperscript{20}. In this process of \textit{adverse selection}, high quality goods are driven out of the market by low quality goods. In addition, the asymmetric distribution of information can also give rise to \textit{moral hazard} in the enforcement of contracts, which means that parties may misuse their information advantage. Examples are painters who use poor quality paint and lawyers who give unfounded advice. The problems of adverse selection and moral hazard may explain the existence of, for example, certificates, licenses and other trading regulations for professional groups such as building contractors, hairdressers and plasterers. By means of these rules, minimum requirements can be set on the commercial knowledge, professional skill and creditworthiness, so that the transaction costs decline and the information problems are reduced\textsuperscript{21}. Because of the nature of credence goods, it is difficult to set minimum quality standards precisely in those cases where the risks of moral hazard are high. In such circumstances, legally sanctioned self-regulation can counter the problems of adverse selection and moral hazard\textsuperscript{22}. Not only do those involved have a vested interest in the maintenance of the minimum quality, they are also better able to formulate and maintain quality rules\textsuperscript{23}.

Problems of adverse selection and moral hazard arise particularly in insurance markets\textsuperscript{24}. Insured parties have superior information available with respect to the incidence of risks but they lack information regarding the quality and inde-

\textsuperscript{19} Nelson (1970); Darby and Karni (1973).
\textsuperscript{20} Akerlof (1970).
\textsuperscript{21} Leland (1979).
\textsuperscript{22} Van den Bergh and Faure (1991); Den Hertog (1993).
\textsuperscript{23} Gehrig and Jost (1995).
\textsuperscript{24} Rothschild and Stiglitz (1976).
dependence of intermediaries. In many countries, social legislation is introduced as a reaction to these problems, and rules are established for intermediaries. Shrinking markets can also arise as a result of search costs incurred by consumers when relevant information is not available to them. Because in their purchases consumers compare the utility of a good with its effective price, search costs give rise to shrinking markets. Search costs can be kept to a minimum through rules related to price and quantity marking, and in the case of credence goods, also for example a ban on advertising\(^{25}\). Finally, under certain circumstances transaction costs can be kept to a minimum by rules relating to misleading information\(^{26}\).

1.2.3.2 External effects and public goods

In addition to information failures, missing markets may also result from prohibitively high transaction costs. Transaction costs can impede, for example, the coming into existence of a market for the efficient use of the environment. In a market economy, resources are efficiently used when the production of goods is increased until the marginal costs equal the marginal benefits of production. In a market with perfect competition, an individual producer aiming for maximization of profit will increase his production until his marginal costs equal his market price. However, an inefficient allocation of resources can arise in the presence of external effects\(^{27}\). External effects are influences of economic action having consequences for the conditions of production or the level of utility of third parties and which come into being outside the market. An often cited example concerns the discharge of waste material by a factory such that for example downstream drinking water companies must incur costs of water purification. Because the private costs for the discharging manufacturer differ from the social costs, production will be increased further than would be desirable from the point of view of efficient allocation. According to the Coase theorem, an efficient allocation of resources can nonetheless result from a process of negotia-


\(^{26}\) Beales, Craswell and Salop (1981); Schwartz and Wilde (1979).

\(^{27}\) Meade (1973).
tion in the case of clearly described property rights and in the absence of information and transaction costs\textsuperscript{28}. The role of the government is to assign property rights to parties involved and allow them to reach private agreements that internalise the externalities\textsuperscript{29}. The information cost and the transaction costs of organization and negotiation can, however, be prohibitively high if several parties are involved in the negotiating process and if an element of strategic behaviour exists\textsuperscript{30}. In such cases, the government can attempt to internalise the external effects by means of regulation\textsuperscript{31}. Examples are safety regulations for items such as automobiles and food, noise levels for aircraft, the obligation to use catalytic converters in automobiles and limits for the emission of hazardous substances in permits.

Next to informational asymmetries and externalities may particular goods with specific characteristics cause markets to fail, the so-called public goods\textsuperscript{32}. Pure public goods have two special characteristics. For the supplier of these types of goods it is first of all either impossible or too expensive to exclude people from consumption who fail to pay for the goods; the technical term for this is non-excludability. In the second place, consumption of these types of goods by one person is not at the expense of the consumption for another person; the technical term for this in non-rivalness\textsuperscript{33}. Classical examples of these types of goods are lighthouses, public order, defence, street lighting and sea defences, and television and radio signals. Public goods are either not produced at all or not in the optimum quantity by a market because of free-rider problems and problems with establishing the ’willingness to pay’ for these goods\textsuperscript{34}. If a supplier has already produced the goods, consumers can be tempted to take a free ride on the willingness to pay of others: after all, they can no longer be excluded from consumption of the good. To establish the optimum quantity of the collective good, the marginal utility of single increments of this good must be known.

\textsuperscript{28} Coase (1960)
\textsuperscript{29} Property rights are interpreted broadly here and include property rules, liability rules and inalienability rules, see Calabresi and Melamed (1972).
\textsuperscript{30} Veljanovski (1982).
\textsuperscript{31} Gruenspecht and Lave (1989).
\textsuperscript{32} Samuelson (1954).
\textsuperscript{33} Musgrave (1969).
\textsuperscript{34} Bohm (1987).
from all the consumers involved. Because of its non-rival character, the aggregate 'willingness to pay' for marginal units is compared against the marginal costs. When consumers are asked to reveal this 'willingness to pay' for extra units, they will exaggerate or minimize this for strategic reasons. Exaggeration will occur when the willingness to pay for extra units is not linked with actual payment for extra units of the good. Minimization will occur when the payment for the public good is linked to the willingness to pay that was indicated. Because consumers cannot be excluded, there will be a tendency to free-ride on others' willingness to pay, and for strategic reasons they will indicate a modest willingness to pay for themselves. For these reasons, a market economy will not be able to produce these goods in optimum quantities, if at all. Government regulation is necessary for establishing the optimum quantity of the goods concerned, and for enforcing the payment of these goods. Many goods, such as education, health care, parks, roads and golf courses have public good dimensions. In such cases also, government regulation can theoretically contribute to an efficient use of resources in an economy.

1.2.4 Undesirable market results

According to public interest theory, regulation can be explained not only by imperfect competition, unbalanced market operation and missing markets, but finally also by the need to prevent or correct undesirable market results. In a competitive market economy, participants in the economic process are rewarded according to their marginal productive contribution. This result of the market process might be undesirable for economic and other reasons. In the first place it may be possible that a redistribution has large incentives effects. In the second place it is possible that an efficient redistribution will increase the general level of economic welfare because dilemma’s such as the prisoner's dilemma impede voluntary transfer. An efficient redistribution could also occur where marginal utility of income diminishes and satisfaction capacity does not differ widely among people. However, in economics it is customary to assume

the unfeasibility of cardinal measurement of utility and interpersonal utility comparison, so that this last form of efficient redistribution cannot be justified from an economic point of view\(^{37}\).

The correction of undesirable market results can furthermore also be considered desirable for other than economic reasons, such as considerations of justice, paternalistic motives or ethical principles. In that case, trade-offs can arise between, for example, *economic efficiency* and equality: incentive effects of redistribution can result in a decline in the level of individual satisfaction of needs\(^{38}\). Public interest theories are usually applied to explain regulation from the pursuit of economic efficiency\(^{39}\). In other cases, the public interest theory is more broadly interpreted and regulation is accounted for as aiming to correct inefficient or inequitable market practices\(^{40}\). According to this last view, regulation can be accounted for as aiming for a *socially efficient* use of scarce resources. Examples of laws and rules intended to prevent or ameliorate undesirable market results are a legal minimum wage, maximum rents, cross-subsidies in postal delivery, telephone calls and passenger transport, rules enhancing the accessibility of health care or services provided by the utilities, rules guaranteeing an income in the event of sickness, unemployment, disablement, old-age etc.

### 1.3 Criticism of the public interest theories

The theory that regulation can be explained as an answer to market failure has been criticized from several points of view.

a. In the first place, criticism has been directed at the *theory of market failure* underlying the explanation of government regulation\(^{41}\). In practice it appears that the market mechanism itself is often able to compensate for any inefficiencies. In that way problems of adverse selection are solved by companies themselves

---

37 Robbins (1932).
38 Okun (1975).
41 Cowen (1988).
by, for example, the issue of guarantees, the use of brand names and extensive advertising campaigns as a signal of quality\textsuperscript{42}. The market sector also appears able to avail itself of goods traditionally characterized as typical public goods, such as lighthouses\textsuperscript{43}. The conclusion that monopoly power or external effects give rise to an inefficient allocation of resources can only be understood by assuming a model in which transaction costs are absent. The allocation of resources appears efficient if transaction costs are included in the analysis\textsuperscript{44}. The assumption of market failure in the case that only one or a few companies are responsible for the production of goods is similarly criticized\textsuperscript{45}. Any significant returns could be a result of superior efficiency of these companies and furthermore and account must be taken of the possibility of competition for the market as opposed to competition in the market\textsuperscript{46}. A more general criticism of the theory of market failure is its limited explanatory power. An economist generally needs only 10 minutes to rationalize government intervention by constructing a form of market failure\textsuperscript{47}.

b. In the second place, the original public interest theories assumed that \textit{government regulation is effective} and can be implemented without great cost\textsuperscript{48}. So precisely the transaction costs and information costs, which underlie market failure, are assumed to be absent in the case of government regulation. This assumption has been criticized in both empirical and theoretical research.

Theoretical research, the theory of the second best, has demonstrated that the partial aim for efficient allocation does not make the economy as a whole more efficient if unavoidable inefficiencies persist elsewhere in the economy\textsuperscript{49}. Unavoidable inefficiencies such as imperfect competition in the commodity markets or the labour markets distort allocation in the whole economy. Not only will the production of a good be produced in insufficient quantities, but also too

\textsuperscript{42} Nelson (1974)
\textsuperscript{43} Coase (1974) ; for externalities, see Cheung (1980).
\textsuperscript{44} Dahlman (1979); Toumanoff (1984).
\textsuperscript{45} Demsetz (1976).
\textsuperscript{46} Baumol et al. (1982).
\textsuperscript{47} Peltzman (1989).
\textsuperscript{48} Posner (1974).
\textsuperscript{49} Ng (1990).
many resources will be devoted to other goods and services in the economy. These distortions also mean that the allocation in the factor market will be sub-optimal. Consider the case that the government wishes to aim for allocative efficiency in an economy in which allocation is sub-optimal. Suppose electricity is a substitute of a product in a market where competition is imperfect. In that case, it is of little use to aim for allocative efficiency in the market for the distribution of electricity through, for example, the regulation of prices of electricity distribution. A 'second best' solution would be to supply electricity at a price lower than marginal costs. In reality however, a great many of these unavoidable inefficiencies exist. These inefficiencies are a consequence of such things as external effects, taxation, imperfect competition and flawed information. That renders the achievement of a second-best optimum unfeasible in practice. Other theoretical research points to the flawed information available to regulators, both in the drawing up of the rules and in their enforcement. A consequence is, for example, that regulated businesses select inefficient combinations of production factors, so that they are unable to produce to minimal costs. Furthermore, x-inefficiencies arise so that production no longer takes place at minimal costs. Another consequence of information problems may be that inefficient (safety) standards are applied by regulators, see for example Viscusi (1992).

Empirical research into the effectiveness and efficiency of government regulation also gives rise to criticism of the public interest theories of regulation. The research into economic regulation was started with the seminal contribution by Stigler and Friedland (1962) about the effects of price regulation on electricity producers. An earlier synthesis of this type of research showed firstly that the influence of regulation on natural monopolies was slight if not non-existent. In the second place, it appeared that regulating potentially competing sectors such

---

50 Utton (1986).
53 Leibenstein (1966).
54 For the effects of regulation, see generally Joskow and Rose (1989).
55 It is ironic that the change in the view of the regulator as being benvolent to the view of the regulator being captured by private interests, was based on a wrong analysis that caused the estimated effect of regulation to be dramatically understated, see Peltzman (1993).
56 Jordan (1972).
as air traffic and freight resulted in an increase in prices and a restricted number of competitors. Empirical research further demonstrates that regulation prescribed an inefficient price structure in which mainly consumer groups received cross-subsidies\textsuperscript{57}. Research into the effects of economic deregulation demonstrated furthermore that mainly consumers, but to some extent also producers, derived a benefit on balance from less government regulation\textsuperscript{58}. Social regulation appeared to keep costs and benefits more or less in balance\textsuperscript{59}, although there is also empirical evidence suggesting that much social regulation is poorly targeted or is over-stringent\textsuperscript{60}. A qualifying remark can be made pertaining to social regulation, that it is hardly if at all possible to monetize or quantify many of the benefits\textsuperscript{61}. Finally, there are arguments for assuming that even competition legislation is misused as an instrument of monopolization and the transferal of rents\textsuperscript{62}.

The points of criticism given under subsections a and b make clear that at the root of the public interest theory lies what is known as the \textit{Nirvana approach}\textsuperscript{63}. Regulation can be explained by assuming that a theoretically efficient institution is able to replace or correct imperfect real institutions.

It is no longer the case that transaction costs and administrative costs under regulatory systems tend to be ignored; see for example the work of what has variously been referred to as the "New Haven" or "Progressive School" of Law and Economics\textsuperscript{64}.

c. Public interest theory usually assumes that regulation can be accounted for as aiming for economic efficiency. Interpreted in this way, the theory is unable to explain why on occasions other objectives such as procedural fairness or re-

\textsuperscript{57} Posner (1971).
\textsuperscript{58} Winston (1993).
\textsuperscript{59} Hahn and Hird (1991).
\textsuperscript{60} Sunstein (1990); Hahn (1996).
\textsuperscript{61} For example, how can a value be put on the preservation of a variety of life forms and how can the preferences of future generations be taken into account?
\textsuperscript{62} Baumol and Ordover (1985); McChesney and Shughart II (1995).
\textsuperscript{63} Demsetz (1968).
\textsuperscript{64} Rose-Ackerman (1988, 1992).
distribution are aimed for at the expense of economic efficiency\textsuperscript{65}. On the other hand, when it is assumed that regulation can be accounted for as aiming for social efficiency, another problem is encountered. Where there is conflict between efficiency and equity, it is hardly possible for at least two reasons to establish the social efficiency of regulation\textsuperscript{66}. In contrast to the case of economic considerations, where the criteria of Pareto and Kaldor-Hicks are generally applied, in the case of considerations of justice there are no generally applicable standards of justice available. No agreement exists regarding the definition of justice in concrete situations\textsuperscript{67}.

Secondly, the establishment of social efficiency of regulation requires that economic efficiency and justice be weighed against each other. A theoretically justifiable and practically usable scale of values that this calls for is not available\textsuperscript{68}. The absence of generally applicable standards of justice and the lack of insight into the relationship between justice and efficiency renders empirical testing of the public interest theory as an explanatory theory of regulation impossible. A key problem of the public interest theory is that the evaluating, normative theory of economic welfare is used a positive explanatory theory of regulation\textsuperscript{69}.

d. A final point of criticism is that public interest theory is incomplete. In the first place, the theory does not indicate how a given view on the public interest translates into legislative actions that maximize economic welfare\textsuperscript{70}. The political decision making process consists of various participants who will aim for their own objectives under different constraints. In contrast to the market economy, it is unclear in the political decision making process how the interaction of the participants will lead to maximum economic welfare.

Secondly, a theory of regulation should be able to predict which branches of industry or sectors should be regulated, and to whom the advantages and disadvantages are to accrue. The theory should also be able to predict what form regu-

\textsuperscript{65} Joskow and Noll (1981), p. 36.
\textsuperscript{66} Sen (1979a; 1979b).
\textsuperscript{67} Dworkin (1981).
\textsuperscript{68} Ng (1985).
\textsuperscript{69} Joskow and Noll (1981).
\textsuperscript{70} Posner (1974).
ation is to take, such as subsidies, restricted entry, or price regulations\textsuperscript{71}. Of course, much normative public interest analysis has been undertaken on the forms of regulation, and not only of economic regulation\textsuperscript{72}. Public interest theory does not appear able to come up with relevant predictions that are amenable to testing by empirical economic science. Furthermore, facts are observed in social reality which are not well accounted for by the public interest theories of regulation. Why should companies support and even aim for regulation intended to cream off excess profits?

1.4 A more sophisticated version of public interest theory

Criticism of the public interest theory has led to the development of a more serious public interest theory of regulation\textsuperscript{73}. According to the naïve public interest theory, regulation can be accounted for by market failure under conformance to the conditions of the Coase theorem. This implies the assumption of absence of transaction costs and freely available, conveniently processed information in the political process. By letting go of these assumptions, a more sophisticated version of the public interest theory comes about. Is it possible to see regulation as an answer to market failure when account is taken of transaction costs and information costs? In the presence of information and transaction costs, regulation can be a more efficient solution to market failure than private negotiations between the parties involved. Costs of organization can also be avoided when for example in the case of environmental pollution, politicians bundle the preferences of those negatively affected. In the case of flawed information, political entrepreneurs can detect the causes of market failure and report them to those involved. In this way, knowledge about for example, accidents can be picked up through safety regulations in factories. Regulation may be more efficient in this case because the government can obtain information less expensively. On the one hand, the government can enforce the provision of information about accidents, for example, and on the other hand information is

\textsuperscript{71} Stigler (1971).
\textsuperscript{73} Noll (1983, 1989).
often a by-product of other government activities. This sophisticated version of
the public interest theory does not therefore require regulation to be perfect. It
does, however, assume that market failure exists, that regulation is the most ef-
fective means of combating it and that regulation does not continue to exist
once the costs exceed the advantages. This theory also assumes that politicians
support open decision making processes and spread information widely about
the effects of market results and regulation. According to this theory, then,
regulation can be accounted for as the efficient solution to market failure. The
problems stated under sections 1.3b and 1.3c are not, however, solved with this
version.

1.5 Private interest theories of regulation

After the public interest theory had fallen into disrepute through empirical and
theoretical research, the capture theory was developed mainly by political scient-
ist$	extsuperscript{74}$. This theory assumes that in the course of time, regulation will come to
serve the interests of the branch of industry involved. For example, it is as-
sumed that legislators subject the branch to additional regulation by an agency
if misuse of the economic position of power is detected. In the course of time,
other political priorities arrive on the agenda and the monitoring of the regula-
tory agency by legislators is relaxed. The agency will tend to avoid conflicts
with the regulated company because it is dependent on this company for its in-
formation. Furthermore, there are career opportunities for the regulators in the
regulated companies. This leads in time to the regulatory agency coming to rep-
resent the interests of the branch involved$	extsuperscript{75}$.

The capture theory is unsatisfactory in a number of respects. In the first place
there is insufficient distinction from the public interest theory, because the cap-
ture theory also assumes that the public interest underlies the start of regula-
tion. In the second place, it is not clear why a branch can succeed in subjecting an
agency to its interests but cannot prevent its coming into existence. In the third

$	extsuperscript{74}$ For a discussion, see Posner (1974).

$	extsuperscript{75}$ For an overview of the various strategies available to be applied by agencies and regulated
companies, see Owen and Braeutigam (1978).
place, regulation often appears to serve the interests of groups of consumers rather than the interests of the branch. Regulated companies are often obliged to extend their services beyond the voluntarily chosen level of service. Examples are transport services, the supply of gas, water and electricity and telecommunication services to consumers living in widely scattered geographical locations. In the fourth place, much regulation, such as environmental regulation, regulation of product safety and labour conditions are opposed by companies because of the negative effect on profitability. Finally, the capture theory is more of a hypothesis than a theory. It does not explain why a branch is able to 'take over' a regulatory agency and why, for example, consumer groups fail to prevent this takeover.

1.5.1 The Chicago theories of regulation

In 1971 a start was made on the development of a theory of regulation called by some the economic theory of regulation\(^{76}\) or the Chicago theory of government\(^{77}\) and by others the private interest theories of regulation\(^{78}\). In that year a seminal contribution by George Stigler appeared, 'The Theory of Economic Regulation'.

1.5.1.1 Stigler

The central proposition of George Stigler was that 'as a rule, regulation is acquired by the industry and is designed and operated primarily for its benefit'. The potential benefits of regulation for a branch of industry are obvious. The government can grant subsidies or ban the entry of competitors to the branch directly so that the level of prices rise and profits increase. In the second place, the government can maintain minimum prices more easily than a cartel. In the third place, the government can suppress the use of substitutes and support complements. An example of support to complements is the subsidization of airports to the benefits

\(^{76}\) Posner (1974).
\(^{77}\) Noll (1989)
\(^{78}\) Baron (1995), who also speaks of the political economy approach to regulation.
of airlines. On the one hand, a demand will therefore arise for government regulation. On the other hand the characteristics of the political decision making process make it possible for sectors of industry to exploit politics for its own ends. For this proposition, Stigler makes use of the insights of Downs (1957) and Olson (1965). In the political decision making process, it will be interest groups who exercise political influence, as opposed to individuals. Individuals will not participate because forming an opinion and producing political action on political issues is expensive in terms of time, energy and money, while the benefits in terms of political influence will be negligible. A representative democracy will more readily honour the strongly felt preferences of majorities and minorities than the less passionately expressed preferences. This is related to the costs of organization of such minority and majority groups. Some groups can organize themselves less expensively than others. Small groups have the advantage because the information and transaction costs are lower and the 'free-rider' problem is smaller than is the case with large groups. Furthermore, in small groups the preferences will be more homogeneous than in large groups. Small groups also have the advantage in that for the same aggregate benefits, the per capita benefit of the group is greater. The fact that apparently large branches can still be well organized is explained by Stigler through concentration and asymmetry\textsuperscript{79}. The large companies in a concentrated industry will see themselves as a small group. In the case of asymmetry in the industry, for example as a result of product diversity or widely varying production techniques, separate companies will wish to prevent unfavourable regulation and will participate in the organization.

The result of variation in the costs of organization is that producers organize more readily than consumers. Not only are the costs more modest for producers, but also the burden of regulation in the form of such things as higher prices per consumer are too slight to justify organization. Politicians aim for re-election. Organized branches can contribute to re-election in two ways: by supplying votes and other resources. Examples of these resources are campaign contributions, chairing fundraising committees and the offer of employment to party members. The larger branches have an advantage in this over smaller

\textsuperscript{79} Stigler (1974).
branches, unless the smaller branches have something in their favour such as a strong geographical concentration. Politicians will honour the demand for regulation by branches because it is not worth while for the majority of opponents to gather information and organize. The conclusion is that regulation is not directed at the correction of market failures, but at setting up income transfers in favour of the industries in exchange for political support.

1.5.1.2 Peltzman and others

In the same issue of the Bell Journal of Economics in which Stigler put forward his theory of economic regulation, Posner implicitly supplied the first criticism\(^{80}\). He observed that in many cases regulation strongly benefited specific consumer groups. For instance, uniform prices were prescribed for such things as rail transport, the supply of gas, water and electricity, telecommunications traffic and mail distribution. The costs of the services supplied differ considerably between consumer groups, however, depending on their geographical spread, among other factors. Other examples are the supply of drinking water to households, schools and fire services, either free of charge or at a price lower than the marginal costs; free rail travel for government workers and military personnel; the supply of electricity to hospitals at less than marginal costs and so on. This phenomenon of internal or cross-subsidization does not fit in with Stigler’s theory of regulation. Even if other consumer groups are obliged to pay higher than marginal costs for their goods and services to compensate, cross-subsidization works against the aim of maximum profit. An explanation of cross-subsidization is provided in an extension to the theory of regulation by Sam Peltzman\(^{81}\). He assumes that politicians will choose their policy of regulation such that political support is maximized. It is not likely that regulation will advantage the industry exclusively. Some consumer groups will also be able to organize themselves effectively. Moreover, organization and information costs are an obstacle to the immediate and total withdrawal of political support in the event of a small decrease in cartel profit. Lower prices are favourable to con-

\(^{80}\) Posner (1971).

\(^{81}\) Peltzman (1976).
sumers, higher prices generate more political support from industry. According to Peltzman, the core problem for regulators is efficient regulation: what price level should be settled on, such that the gain in votes resulting from the income transfer just balances the loss of votes resulting from the rise in prices. This extended theory explains not only the phenomenon of cross-subsidization, but also predicts which sectors or industries will be regulated. These are the relatively competitive industries and the monopolistic branches. In the first case, the industries have a keen interest in regulation and, in the second case, consumers have a great interest in regulation. The intermediate industries will expect that any regulated price level will not deviate widely from the actually existing price level. In that case it is not worthwhile for either consumers or producers to organize to acquire favourable regulation. The practice of regulation appears to confirm this prediction. Regulated branches are either monopolistic, such as rail transport and telecommunications, or highly competitive, such as freight, agriculture, independent professions and cab companies.

As well as the types of branches, the theory of regulation also predicts the form the transferred benefits will take. In principle, transfers can come about directly through subsidies or indirectly through price or quantity regulation or restriction to market entry. Stigler originally assumed that branches would choose indirect support. The granting of subsidies would result in entry, so that the subsidy per producer would be dissipated. In an extension to the theory of regulation, Migué has shown that the form of the transfers is partly dependent on the elasticity of supply of the production factors in the branch concerned. In the political market, the public are both consumers and suppliers of production factors. Suppliers of production factors will give preference to subsidies when the supply is inelastic. The taxation necessitated by the subsidy is distributed over many tax payers, while the subsidy accrues to a limited group of suppliers of production factors. This extension to the theory of regulation explains why subsidies are granted in sectors such as education, health care, domestic housing, and city transport and why quota systems and price regulation can be found in sectors such as agriculture, airlines, road transport and railways. Similar reasoning explains why polluting companies give preference to prescribed

---

82 Migué (1977).
soning explains why polluting companies give preference to prescribed limitation of production (quotas) above taxation.\(^{83}\)

Another extension to the theory of regulation is from McChesney.\(^{84}\) He sees politicians not as neutral agents brokering between competing private interests directed to obtaining transfers of income. In his view, politicians also try to gain advantages by putting private parties under pressure. He gives examples in which Congress, under the threat of price reductions or cost increases, forces concessions from private parties. To make such rent extractions possible, politicians encourage private parties to organize. Organization thus not only enhances the probability of gaining transfers of income, it also increases the risk of having one’s own surplus threatened and expropriated.

Finally, Keeler has supplemented Peltzman’s model with public interest considerations.\(^{85}\) In his model, politicians not only gain political support through transfers of income between interest groups. By increasing economic efficiency, for example in the case of economies of scale and externalities, resources are acquired which can be distributed among producers and consumers. Rational politicians will take advantage of this possibility.

1.5.1.3 Becker

A further contribution to the Chicago theory of regulation was made by Gary Becker.\(^{86}\) He concentrated on the consequences of the competition between interest groups, which he calls pressure groups. As the political pressure increases, political influence also increases and the financial benefits from the pressure exerted rises. Some groups are more efficient in the exertion of political pressure than others. This can be a consequence of the economies of scale in the production of pressure, more effective combating of free-riding, better access to the media and other matters. In this way, wealth is transferred from less efficient to

\(^{83}\) Buchanan and Tullock (1975).
more efficient groups, in the form of subsidies, but also through such things as price regulation. A limit exists for these transfers, however. The transfers provoke economic welfare losses, which are known as the deadweight costs. As a result of these welfare losses the loss to the less efficient pressure group is larger than the gain of the more efficient pressure group. As the welfare losses increase, the pressure of the more efficient group will decline because the benefits of pressure are lower. At the same time, the pressure of the less efficient group increases with the magnitude of the welfare losses because the potential gain of pressure increases. This countervailing pressure limits the possibility of transfers to the more efficient pressure group. It can be deduced from this analysis that politically successful groups are small in proportion to the group bearing the burden of the transfers. The larger the burdened group, the smaller the transfer per member of the disadvantaged group and the smaller the deadweight costs. This diminishes the countervailing pressure. The smaller the receiving group, the larger the potential gain per member of the group, which serves to increase the pressure exerted. The analysis explains for instance, why in countries where the agriculture sector is small it is subsidized, while large agriculture sectors elsewhere are heavily taxed.

In Stigler's and Peltzman's view, competing branches have information and organization benefits through which favourable regulation can be predicted. In practice, such regulation of competitive sectors is rarely seen. The explanation is found in Becker's theory of competition among pressure groups. Losses of welfare are greater where the elasticity of supply is greater. In competitive sectors, these elasticities are large. The transfers of income and the welfare losses associated with regulation are so large that the countervailing pressure invoked eliminates the investment in political influence.

It can be further deduced from the analysis that regulation is more likely of economic sectors exhibiting market failures, a result in agreement with the public interest theories of regulation. In monopolistic industries, for example, consumers can obtain transfers larger than the accompanying losses to the producers. In competitive sectors, on the other hand, the gain of the winners is smaller than the total loss of the losers. All other things being equal, more pressure by the potential winners will be exerted in monopolistic sectors and less pressure by the potential losers than in the competitive sectors. Market failure is therefore not a sufficient condition for regulation, as in the public interest theories of
regulation; regulation also depends on the relative efficiency of pressure groups in exerting political pressure. In contrast to Olson (1982), the competition between pressure groups will not have any negative effects on the growth of the national product and productivity, at least provided pressure groups of equal size and efficiency are involved in producing the pressure. The competition between pressure groups will also lead to the most efficient form of regulation. Even if under certain circumstances the results of competition among pressure groups is efficient, Becker claims that production of pressure is not. All pressure groups would be better off if they decreased their expenditure on pressure by equal amounts. Various laws and rules directed to limiting the influence of pressure groups can be explained as instruments for opposing wasteful expenditure on political pressure. The Chicago theories of regulation seem primarily suited to the explanation of so-called economic regulation. Social regulation, the regulation in the area of safety, environment and health, seems at first sight to be less amenable to explanation by these theories. There are diseconomies in the area of organization, the benefits are divided among many involved parties and the costs of regulation are allocated to concentrated groups. Nonetheless, private interest explanations have also been put forward for some examples of social regulation, see for example Bartel and Thomas87 and Pashigian88. There are a number of ways how private interests may benefit from social regulation. First, the implementation of rules and standards will be in the interests of those companies already complying with the standard. Furthermore, large companies have an advantage when it is necessary to comply with administrative obligations or costly measures that have the nature of sunk costs. Small companies may be driven out of the market, so that the competition is limited. Legal obligations are often differentiated into existing producers and new producers. By setting higher standards on new producers, entry to the market is impeded and competition is limited. Finally, if regulation requires provisions and facilities, the producers of these provisions and facilities will benefit from regulation.

87 Bartel and Thomas (1985, 1987)
1.6 Criticism of the Chicago theories of regulation

A weak point of the Chicago theory of regulation is its risk of becoming a tautology\textsuperscript{89}. Redistribution is seen as the cause of regulation. In practice, however, regulation is always associated with redistribution. By investigating who derives the benefits from regulation and who carries the costs, one has not established the cause of regulation. Another weak point is that it cannot be predicted which groups will be the most effective politically and who will collect the transfers of income. Research has shown that it was workers who enjoyed the main benefits of regulation rather than producers. This result might be rationalized but is difficult to predict from the Chicago theories. Furthermore, the Chicago theories are incomplete\textsuperscript{90}. The Chicago theories of regulation assume that interest groups determine the outcomes of elections, that legislators honour the wishes of the interest groups and that are no principal-agent problems between legislators and regulators. In this theory of regulation there is therefore little or no attention to the following three subjects:

a. the motivation and behaviour of the various political actors, such as voters, congressmen, legislators, bureaucrats and agencies;
b. the interactions between the various actors in the regulation process;
c. the mechanism through which legislators and regulators conform to the wishes of the organized private interests.

In fact it is assumed that the operation of the political process of legislation and regulation has hardly any independent influence on the pattern and form of regulation, if at all. This assumption has been criticized from several quarters and several attempts have been made in the literature to fill in the three gaps. In these developments, the theory of regulation partly overlaps with the public choice theory\textsuperscript{91}.

\textsuperscript{89} Noll (1989a).
The idea that only organized interest groups are being favoured by legislators has been criticized by Wilson (1974, 1980). The origin of regulatory legislation can, according to Wilson, be explained by analysing the concentration and dispersion of costs and benefits. Majoritarian politics in Congress is to be expected with both widely distributed costs and benefits; antimonopoly legislation is one example. Interest group politics arise with both concentrated costs and concentrated benefits; labour legislation and railway regulation are examples of this. Client politics is the result of concentrated benefits and widely dispersed costs; the protection of a number of professions by means of licensing is considered to be an example as is the subsidization of companies and industries. A final form of policy is what is known as entrepreneurial politics, in which the costs are concentrated and the benefits distributed; examples are the protection of the environment, of consumers against unsafe products and of workers against industrial accidents and occupational illnesses. This last form of regulation is difficult to bring in line with the Chicago theories of regulation. According to Wilson, interest groups are therefore not the only origin of regulation. Wilson goes on to criticize the assumption of the Chicago economists that legislators are able to control regulators unimpaired. In his view, the behaviour of the agencies can be better accounted for through an analysis of the motivations and restrictions of those involved internally. He distinguishes careerists, professionals and politicians and uses this to account for various types of regulation policy. Price regulation, for example, will be simple in structure for careerists and more complex for professionals.

Contrary to Wilson, Derthick and Quirk assume that the regulatory strategies followed by the agencies is actually heavily influenced by surrounding forces other than concentrated or diffused benefits and costs. In their study they account for deregulation by pointing to the importance of the intellectual climate in combination with the pressures exerted by the president, Congress and the judiciary on the agencies. In agreement with the Chicago theories of regulation, Weingast also sees no independent role for agencies. Changes in the regu-

---

*regulation*. No general theory has emerged however, there are mainly a number of separate contributions to the field by public choice scholars.

92 See particularly his study of governmental agencies (Wilson, 1989).

93 Derthick and Quirk (1985).

tory behaviour of agencies are a consequence of changes in the preferences of Congress or its commissions. Contrary to Chicago however, Weingast shows how a structure-induced equilibrium of policy choice arises as a reaction to the instability of majority rule voting. In an equilibrium of congressional committees, the agencies and the interest groups have divergent but comparable goals. Weingast's model describes why the objectives of interest groups are honoured and how diffuse single-issue groups such as environmental groups and consumer organizations are able to acquire political power at the expense of traditional interest groups such as the industry, the employees and agriculture.

The assumption that legislators consistently act in line with the interests of the interest groups comes in for particular criticism from the point of view of the principal-agent theory. The acquisition of information about the behaviour of legislators and regulators is expensive. Because of this, legislators and regulators may escape the attention of voters and interest groups and act in their own interests or according to their ideological preferences. In this connection, Levine and Forrence distinguish two types of motivation and two types of political dominance. Depending on what a political actor is aiming for, private and public interests can be distinguished. Private interests are preferences of political actors with respect to their self-interest. Public interests are preferences related to the interests of others. The private and public interests indicate what a political actor will maximize when there is room to aim for their own preferences. General-interest policy is a policy that should be ratified in the absence of information, organization, transaction and monitoring costs. Special-interest policies should not be ratified by the general polity in the absence of monitoring costs and transaction costs etc. Because of the existence of these costs, a ‘slack’ or policy drift arises, in other words agents have room to pursue their own objectives. On the one hand this discretion can be used to favour special interest groups, on the other hand, the policy slack can be used to promote the interests of others. Here again there are two possibilities. In the absence of monitoring costs etc., the regulatory policies would either not be ratified by the general polity or it would be.

---

95 For example, Kalt and Zupan (1984,1990).
The capture debate is concerned with the question who dominates the political process and is therefore concerned less with private or public interests than with general and special interests. According to Levine and Forrence, general interests will prevail to the extent that the slack is reduced. The amount of slack decreases drastically when issues come up on the public agenda. Favourable conditions for this are: political competition, special interest organizations, public policy intelligentsia and the news media. A general-interest policy does not otherwise imply that the policy is also efficient. When for example the rented sector is a substantial part of the housing market, a policy of rent control will achieve ready approval without any guarantee of efficiency.

A conceptually different criticism of the Chicago theories of regulation comes from the Virginia School of Public Choice (Rowley, Tullock, Tollison, McCormick et al.)\(^{97}\). In their theories, the term coined by Ann Krueger (1974) *rent seeking*, is a central feature. Rent seeking means the political activity of individuals and groups to devote scarce resources to the pursuit of monopoly rights granted by governments. The Virginia School criticizes the Chicago theoreticians for their disregard of the inefficiencies of regulation. With their emphasis on the inefficiencies of rent seeking, the Virginia School practices mainly normative economic theory; behaviour and institutions are judged according to the degree of efficiency in the allocation of scarce resources. In what has become a classic contribution, Tullock\(^{98}\) has shown that the inefficiencies of monopoly consists not only of the welfare losses known as the *Harberger triangle*, but also of the resource costs emanating from the competition among potential monopolists to acquire monopoly rights. Furthermore, incumbent monopolists will invest resources in maintaining monopoly rights. Graphically, the inefficiencies measured by the Harberger triangle, should be augmented with the *Tullock rectangle*, the maximum monopolists are willing to invest to acquire and maintain monopoly rights which is measured by the expected amount of profits.

\(^{97}\) For an overview see Tullock (1993), Buchanan, Tollison and Tullock (eds.) (1980) and Rowley, Tollison and Tullock (eds.) (1988).

\(^{98}\) Tullock (1967).
Furthermore, the potentially disadvantaged consumers will also apply scarce resources to prevent the creation of a monopoly if possible. After the creation of a monopoly, scarce resources will be wasted because the monopolist will protect his monopoly rights against possible threats, from potential competitors and disadvantaged consumers. Finally, monopoly rights can cause x-inefficiencies: the monopolist will not produce at minimum cost of production\textsuperscript{99}.

In Stigler's and Peltzman's views a non-contestable monopolist will not aim for regulation because no higher profits can be achieved if the industry is regulated. However, according to the Virginians, the incentives to regulate remain, though now from the side of the bureaucracy and politics. The interests of bureaucrats and politicians can be served through giving certain groups of consumers the privilege of cross-subsidies in a disguised form\textsuperscript{100}. Furthermore, according to the Chicago economists, redistributive instruments such as taxation, subsidies or regulation are equivalent either with respect to efficiency or to the precise nature of political the equilibrium. The Virginians point, however, to the visibility of taxes and subsidies and the waste of scarce resources through the reactions provoked by such instruments. Regulation gives more room to politicians and bureaucrats to put their own objectives into effect. Between regulation and taxation there are also large differences concerning the degree of inefficiency. Traditionally, account is taken only of welfare losses in the form of the Harberger triangle, which under certain conditions may actually be smaller in the case of taxation than when regulation is applied. All else equal, Chicago economists predict the use of taxation as the preferred instrument to transfer resources. It is to be expected however, that the Tullock rectangle inefficiencies will be larger with taxation than with regulation.

The rent seeking theorems have been criticized for the overestimation of the assumed losses of welfare. It is not likely that monopolists are forced to use the entire Tullock rectangle in order to acquire their monopoly and furthermore, rent seeking outgoings also have positive effects on welfare\textsuperscript{101}. A conceptual

\textsuperscript{99} As a result of x-inefficiencies (Leibenstein, 1966); however on the existence of x-inefficiencies, see also Stigler (1976).

\textsuperscript{100} Crew and Rowley (1988).

\textsuperscript{101} Varian (1989).
and methodological criticism is given by Samuels and Mercuro (1992), who judge that the restrictive assumptions of the rent-seeking school will lead to misleading conclusions. Furthermore, the normative analyses are too selective to serve as a basis for policy.

1.7 Regulation, deregulation and re-regulation

Once the Chicago theories of regulation had been developed, social developments seemed to refute it. While this theory explained regulation as aiming for transfers of income and wealth, at the end of the seventies and the beginning of the eighties, many complexes of rules were dispensed with in a process of deregulation. This deregulation was mainly concerned with economic regulation of sectors such as transport (airlines and freight), telecommunication, energy and the financial sector. The social regulation increased in scale, even though the nature of this regulation sometimes changed (more cost-benefit analyses, more risk analyses, more performance standards, fewer specification standards)\textsuperscript{102}. From the theories of regulation discussed here, various explanations can be derived for this process of deregulation\textsuperscript{103}. From the public interest theories two explanations of deregulation can be derived. In the first place, it is possible that the cause of market failure is removed by technological or demand factors. Through a strongly increasing demand for, for example, transport facilities, a natural monopoly can change into a competitive market. Furthermore, technological developments such as communication via satellite instead of by cable can undermine natural monopolies. A second explanation for deregulation is the availability of more efficient alternatives than regulation to solve market failures. New instruments may have been developed such as auctioning or


\textsuperscript{103} Den Hertog (1996); Keeler (1984); Peltzman (1989).
yardstick competition. It is also possible that gradually more knowledge accumulated and disseminated on the envisaged and real effects of regulations. Finally, it is possible that theoretical developments, such as, for example, the theory of contestable markets, inspire more confidence in the operation of the market mechanism.

At least four driving forces of deregulation can be derived from the *Chicago theories of regulation.* In the first place, shifts can come about in the relative political power of pressure groups, for example, as a result of the more efficient combating of free-riding, the more efficient use of media or as a result of special entrepreneurship (Ralph Nader). In the second place, deregulation can arise when politically effective groups believe that they can better promote their economic interests in an unregulated market, for example by self-regulation. In the third place, deregulation may result from declining profits and the accompanying decline in political benefits of regulation. The fixing of prices or the introduction of entry restrictions in principally competitive sectors, such as the airlines or freight, will result in competition taking place in other dimensions of the product. Competition in the area of service, such as the frequency of transport, will result in a decline in profits. In Becker's view, that leads to a decreased pressure from the industry involved and an increased pressure from consumers for price reduction. In Peltzman's view, politicians will seek other opportunities for regulation with higher political rewards. Finally, deregulation can be accounted for by increasing deadweight costs (Becker). These costs increase in the course of time because substitutes for regulated products are developed and because costly methods of evading and avoiding particular regulations are discovered. The deregulation of sectors such as transport, telecommunications and banking, can then be seen as an echo of the regulation movement of the thirties. Increasing deadweight costs are in the second place a result of the increasing marginal tax rates in the sixties and seventies. According to Becker, this stimulated the pressure on tax payers who were able to collect more political support than the groups who had the benefit of social security programs.

---

105 Baldwin (1990); Stewart (1985); Wilson (1984); Wolf (1979).
According to the public choice theories of regulation\textsuperscript{107} deregulation can first of all be accounted for by a changed balance of power of pressure groups. In the second place, a structure-induced equilibrium can be disturbed by the actions of political entrepreneurs, such as the chairpersons of regulatory commissions. In the third place, politicians can seek political support for deregulation by providing voters with information about the inefficiencies of regulation. Alternatively, politicians could try to use the complexities of regulatory issues by claiming that economic deregulation would greatly advance economic and social welfare.

A general comparison of deregulation practices and the various (de)regulatory theories gives a mixed picture, see Peltzman (1989) and Noll (1989b). If the public interest theories were generally applicable, deregulation would have taken place sooner. On the other hand, events such as the deregulation of freight are difficult to account for with the Chicago theories of regulation. In that sector huge profits were being made and rents were shared with the employees who consequently also had much to lose from deregulation. Various circumstances including political entrepreneurship were considered to be applicable and did in fact influence the practice of deregulation. In other cases, Congress has played no role and the legislation was changed after deregulation was already a fact.

Also the expectations with respect to the future development of regulation and deregulation are mixed. On the one hand there are also researchers who are convinced of the relative efficiency of the market mechanism and of legal foundations that support and sustain the market\textsuperscript{108}. They see a greater role for government in the area of competition politics and in setting constraints to the functioning of markets, such as in the area of safety. On the other hand, there are voices who assume that the process of deregulation will be followed in the downward phase of the business cycle by a phase of renewed regulation\textsuperscript{109}. Possible disadvantages of deregulation, such as predatory pricing, fluctuating prices and discriminatory prices, insufficient service, increased lack of safety,

\textsuperscript{107} Supra, footnotes 91-97 and accompanying text.
\textsuperscript{108} Kahn (1990); Hahn (1990).
\textsuperscript{109} Cudahy (1993).
job insecurity and redundancy for large groups of employees may ring in a new age of regulation.

1.8 Summary

This article makes a distinction between types of theories of regulation: public interest theories and private interest theories\textsuperscript{110}. The Chicago theories of regulation, the theories of regulation by the Chicago economists, notably Stigler, Peltzman and Becker, are the foundation and the constituent part of the private interest theories of regulation.

The Chicago theories are mainly directed at the explanation of economic regulation; public interest theories and public choice theories envisage in addition to that an account of social regulation. The core of the different theories is discussed as well as the criticisms that have been levelled at them. The aim is to infer from the theories in what sectors regulation may be expected and what form the regulation will take. The extent to which these theories are also able to account for deregulation, and the expectations for the future, are discussed.

Acknowledgement

The author is grateful for the valuable comments made by Anthony Ogus and two anonymous referees, none of whom may be held responsible for the views expressed in this paper.

\textsuperscript{110} For some authors, for example Ogus (1994), private interests include the interests of such groups as bureaucrats, administrators or politicians who pursue their own interests. According to this last interpretation the private interest theories may be distinguished in the public choice theories of regulation and the Chicago theories of regulation.
CHAPTER 2 THE EXPLANATION OF DEREGULATION IN THE NETHERLANDS

2.1 Aims and scope of the paper

In the Netherlands deregulation has been on the political agenda since 1982. In that year deregulation became part of the coalition agreement, which is an agreement between the political parties constituting the government. Since then the process of deregulation has not been very successful. However, at the beginning of the nineties deregulation became once again a political priority. This time the process of deregulation appears to have been a success. Why was deregulation a failure at the beginning of the eighties and ten years later a success? To answer this question I will, in the second part of my paper, discuss three theories of regulation. Since regulation logically (and historically) precedes deregulation these theories should be helpful in identifying the relevant factors causing deregulation. In the third section I will describe the politics of deregulation in the Netherlands and the sectors of the economy that were to be deregulated. The fourth section evaluates the process and results of deregulation against the relevant elements of the three theories of regulation. This section leads to some conclusions concerning the causes of regulatory failure and success of the politics of deregulation. The paper will be concluded with a summary of the main results.

2.2 Three theories explaining deregulation

In the 'Law and Economics' literature one can distinguish three strands of regulation theory: the public interest theory of regulation, the Chicago theory of regulation and the public choice theory of regulation\(^1\). I shall give a short

\(^1\) Sometimes the public choice theories and the Chicago theories of regulation are taken together under the heading of ‘private interest theories’ (Ogus, 1994); more often private interest theories of regulation solely refer to the Chicago theories of regulation.
Chapter 2

summary of each of these theories and describe what, according to these theories, are the relevant factors explaining deregulation.

2.2.1 The public interest theory of regulation

The public interest theory takes market failures as a point of departure. A market failure occurs when the unregulated market outcomes are sub-optimal relative to some social welfare function. Standard examples of such markets failures are natural monopoly, externalities, public goods, information failures and the instability of a market equilibrium. In all these instances there is discrepancy between prices and marginal (social) cost so that in principle all market participants can be rendered better off by remedying the market failure. A market failure is of course not a sufficient reason to animate government regulation. Generally speaking, there are several public and private ways of dealing with market failures. For instance, in the case of negative externalities, alternatives to regulation are the creation of a market through marketable permits or the use of taxes and subsidies; in theory, a natural monopoly could produce a competitive market outcome either by the use of public franchising or by the use of such instruments as 'yardstick competition'. Therefore, not only does the public interest theory require a market failure but also that regulation is the most efficient way of bridging the gap between marginal social benefit (price) and marginal social cost. The naïve version of the public interest theory presupposes perfect information and perfect political competition: the very transaction costs that cause the market failure are absent in the political process. According to a more sophisticated version of the public interest theory, regulation might still be the preferred alternative even if transaction costs are present. Economic actors will use the political process if it is less costly than private negotiations and if the welfare losses of the market failure are higher than the cost of regulation. The reason that the political process might be the cheaper way of solving the market failure problem is the coercive power of the state. This power makes it possible to obtain the necessary information more cheaply

\[2 \text{ Noll (1989).}\]
and makes it also possible to effectively oppose free rider problems and strategic behaviour. Implicit in the sophisticated version of the public interest theory is the presumption that such a thing as ‘the public interest’ exists, that government officials act according to the public interest and that the separation of policy making and policy implementation has no bearing on the maximization of efficiency.

From the point of view of the public interest theory of regulation, two major motives for deregulation can be identified. Firstly, demand and technological developments may have eliminated the market failure. Secondly, regulation has turned out not to be or is no longer the most efficient way of dealing with the market failure.

2.2.2 The Chicago theory of regulation

According to the Chicago theory of regulation, or as it is sometimes called, the economic theory of regulation\(^3\), it is not social welfare which a politician maximizes. Politicians and their constituents are rational individuals who maximize utility in an institutional context that defines constraints and opportunities. Politicians are supposed to be motivated by the income, power and prestige derived from being in office. To further these aims, politicians will supply policies, such as price and entry regulations, which maximizes the number of voters. For the individual constituent it is rational to be 'ignorant' and not to take political action. It is therefore, typically, groups that can deliver the necessary resources (votes or campaign contributions) for the re-election of politicians. But groups differ in the likelihood of formation and the relative political effectiveness due to differences in information and organization cost. According to Stigler these differences in cost will tend to create favourable regulation for producer groups\(^4\). Consumers will find it more costly to organize because of their number and the incentives to free ride. Furthermore, a given price increase will result in a far larger per capita stake for producers compared

\(^3\) Posner (1974).
\(^4\) Stigler (1971).
with that for consumers. More generally, the smaller ‘single issue’ groups will tend to dominate the political process at the expense of the larger, diffuse groups, such as consumers, taxpayers, the unemployed and the poor. Peltzman generalized the Chicago theory of regulation by suggesting that politicians maximize total political support by allocating the benefits of regulation across all politically effective groups\(^5\). For instance, politicians will regulate the prices of monopolies downward if the reduction in support from the producers is less than the offsetting growth in support from consumers. Politicians will seek to preserve a politically optimal redistribution of income, once it has been established. Furthermore, the benefits of the reduction in producer income need not be spread uniformly across all consumers. If consumers have equal political weight but differ in the cost of serving them, there will be a tendency to cross-subsidize the high-cost consumer. Well-known examples are short-distance telecommunications and railway and airline services to small communities. The redistribution of income caused by regulation is accompanied by a deadweight loss. The deadweight losses grow larger the more output moves away from the competitive level. But since the political pressures from gainers and losers are approximately proportional to the size of their gain and loss one would expect only the least inefficient regulations to survive. This is an important result derived from Becker's analysis on the competition among pressure groups\(^6\). Another insight from his analysis is that policies that increase efficiency are more likely to be adopted than those that decrease efficiency. The potential increase in aggregate income will induce greater pressure from gainers and less opposition from the losers.

What according to the Chicago theory of regulation are the circumstances under which deregulation is likely to occur? Firstly, changes in the cost of information and organization could lead to effective political representation for the victims of regulation. For instance, the incentives to free ride in the formation of a consumer organization could be checked by providing goods or services to members only. Secondly, deregulation will occur if the current politically effective groups believe they can further their economic interest better in an

\(^5\) Peltzman (1976).
\(^6\) Becker (1983).
unregulated market, for instance by way of self-regulation. Thirdly, a decline in the industry's profits, for instance as a result of lower demand or higher input prices, upsets the political balance of pressures which may result in deregulation. Finally, if deadweight losses increase over time the bias in favour of the status quo inherent in Becker's analysis may be undermined by external pressures.

### 2.2.3 The public choice theories of regulation

The theories of regulation discussed so far treated the political process as a black box, presuming that its workings have no influence on regulatory outcomes. The public choice theories of regulation on the other hand explicitly analyse the behaviour of the different actors involved in the regulatory decision-making process and their interaction. In some theories the emphasis is on the internal structure and workings of, for instance, congress or the agencies involved, others concentrate on the interaction of congress, voters, interest groups and agencies, often in terms of principal-agent theories. The former group of theories, the *structuralist* theories, explains the institutional structure of the legislature from the instability of majority rule voting. Constraints on policy change such as the congressional committee system and rules of policy choice have been developed to counterbalance this instability. Congressional committees have political jurisdiction over certain policy areas, that is, they have special powers such as control over proposals for new legislation, agenda control and agency policy oversight. Since the committee system for the important policy areas is paralleled by the structure of interest groups, so called 'iron triangles' arise consisting of committees, agencies and interest groups who have divergent but comparable goals. A structure-induced equilibrium of policy choice will develop which remains stable as long as the relevant variables do not change. Changes in these variables such as presidential initiative, precedential court decisions or changes in public opinion and balance of power of interest groups may upset the status quo. However, existing relationships

---

Chapter 2

will only be given up by the committee members when the expected political returns from this change are higher than the expected losses. This explains why, initially, a policy will remain unchanged despite changes in the underlying variables.

A second group of theories, the *behavioural* theories, although related to the first, focuses on incentives and the behaviour of political actors rather than on the political structure. The aim is to describe the mechanisms by which voter's preferences are transformed by the political process into policy. Information costs and the costs of monitoring and controlling the behaviour of legislators, committee members, agency heads and their bureaucracies create 'slack' which allows policy discretion for the political actors. This policy discretion can be used by the political actors to favour special interest groups, to further their own conception of the public good (ideology) or to pursue policies that are according to the general interest\(^8\). The amount of slack is influenced by such factors as the degree of political competition, the existence of interest groups and the opinion of academic scholars and journalists. Politicians may try to use or reduce the amount of slack. They may seek political support by providing the public with free information on important issues or by trying to change the issues on the public agenda. Alternatively, politicians may use slack to pursue special-interest policies.

How can these theories help to explain deregulation? *Firstly*, a change in the balance of interest groups could result in deregulation. For instance, since the mid-1960s there has been a rise in single-issue groups such as consumer, environmental and tax limitation groups in the US. This change has been paralleled by changes in the parliamentary committee system which resulted in a decline of political support for the status quo. Furthermore, presidential appointments to regulatory commissions with broad regulatory discretion made de facto deregulation possible without changing legislation. *Secondly*, the status quo could be disrupted by political entrepreneurs, such as the chairmen of the regulatory commissions. According to some scholars, these chairmen

---

\(^8\) Levine and Forrence (1990).
were pushed in the direction of deregulation by prevailing ideas which originated from the critical writings of economists on the causes and effects of regulation\textsuperscript{9}. \textit{Thirdly}, politicians may successfully seek political support for deregulation by providing the voters with information on the inefficiencies of regulation. Alternatively, politicians could try to use the complexities of regulatory issues by claiming that economic deregulation would greatly advance economic and social welfare.  

In order to evaluate which of these theories of regulation has the best explanatory power of deregulation in the Netherlands I shall, in the next section, give a description of the organization and the results of the policy of deregulation.

\subsection{2.3 Deregulation in the Netherlands}

Deregulation became part of the official government policy in 1982 when it was included in the coalition agreement\textsuperscript{10}. The aim of deregulation was primarily the strengthening of the market sector of the economy. However, the precise objective of the policy of deregulation remained, to a large extent, unclear. The objectives of deregulation were both defined as the abolition and reduction of regulation and as the simplification of legislation and regulation. Contrary to the US, where deregulation involves the rule-making powers of specialized government agencies, in the Netherlands deregulation involves regulation and legislation of the central government.  

To achieve the aforementioned objectives, two committees were set up; one was given the task of formulating instructions on the drafting of new legislation and regulation and of making recommendations for the reduction and simplification of areas of existing or proposed legislation and regulation. The other committee had as its specific task to make recommendations on the reduction and simplification of areas of legislation that were important for the operation of the market economy. Examples of such areas of regulation are competition

---

\textsuperscript{9} Derthick and Quirk (1985).

\textsuperscript{10} Deregulation was one of the government programmes referred to as 'comprehensive operations'; other such operations were 'Privatization' and 'Decentralization'.

47
law and regulation, establishment regulation, price regulation, the regulation of shop opening hours, the regulation of sales, work place regulation and haulage regulation.

The members of the committees were mainly top officials in the government service, except for their chairmen who were independent experts (professors) on the matter of administrative law. Two of the eight members of the committee on economic legislation were members on the board of multi-national corporations. The recommendations of the committees were based on reports prepared by the bureaus of the Ministries that implemented the legislation and regulation concerned. The final responsibility for the policy of deregulation was with the Ministry of Justice.

The policy of deregulation is generally considered to be a failure\textsuperscript{11}. Government instructions on the drafting of new legislation and regulation have been issued and all proposed legislation must pass a so-called ‘deregulation test’. Proposed legislation has to be tested against fifteen criteria of ‘good’ legislation such as proportionality, minimum costs of compliance, effective enforcement and the like. However, the deregulation test is being passed without exception. The committee on the reduction and simplification in general did advise on several areas of regulation but the recommendations mainly concerned technical improvements of legislation; an exception was the recommendation to abolish the capacity regulation of road haulage. There were few suggestions on substantial reduction of regulation. The committee on the reduction of economic legislation did not recommend any significant reduction of economic regulation.

This does not mean that the process of deregulation came to an end. At the beginning of the nineties, deregulation became once more a political priority. As a result, a reform of the competition policy regime will soon be implemented. An entirely new legislation will bring Dutch competition policy in line with the European model. Current competition legislation, based on what are referred to as abuse principles, will be replaced by a prohibitory system which means that all major cartels will be prohibited. Also a liberalisation of the establishment laws will be enacted. At this time there are

\textsuperscript{11} Geelhoed (1986); De Ru (1991).
different establishment regulations for 90 sectors of the economy comprising about half the total numbers of enterprises in the Netherlands. In each case, a special licence prescribing certain minimum standards of quality is required to undertake an activity in such a sector. After the liberalisation and harmonization there will be only general requirements for 40 sectors of the economy, no licence will be required for 30 sectors of the economy and special standards will be required for about 20 sectors of the economy.

In the coalition agreement of 1994 a new institutional design of the process of deregulation was announced. The project is called Market Operation, Deregulation and Quality of Legislation. A ministerial committee presided over by the Prime Minister will prepare decisions to be made by the cabinet. The other members of the ministerial committee are the Minister of Justice, the Minister of Economic Affairs and the Minister of Finance. In 1995 working groups of experts and civil servants presided over by an independent chairman have evaluated the Shop Opening Act, Taxi Regulations, Environmental Regulations, Driving Time Regulations, Workplace Regulations and the Process Monopoly of Lawyers. The conclusions of the cabinet have been sent to the parliament for decision making. According to those conclusions, the taxi market will be deregulated, the number of licences required for environmental purposes will be halved through the simplification of regulations and shop opening hours will be extended. Furthermore, legal experts from, for instance, the organization of consumers or the labour organization will be allowed to represent their members in juridical procedures. More flexibility in the planning of driving time and resting hours of drivers will be allowed and performance standards instead of specification standards will be prescribed regarding safety in the workplace. Since these proposals are hardly disputed in parliament one generally expects them to be adopted.

A new round of evaluations has been announced: organizations of employers, labour organizations, the organization of consumers and the Departments have been asked to list the regulations that, according to them, are the most harmful. Finally, proposed legislation will, in the future, be tested against criteria of ‘good’ regulation (deregulation test), against the effects of regulation for the operation of a market economy (market operation test) and against the effects on the environment (environment test). Government bureaus that prepare
Chapter 2

legislation are also obliged to compare different solutions to the same problems from other countries (foreign countries test).

2.4 Causes of successes and failures of deregulation

What we need to explain are the following events. First, how did deregulation become an issue on the political agenda and why was the policy of deregulation in the eighties not a success? Secondly, an explanation is needed for the successful deregulation of the establishment regulation and the re-regulation of the competition policy. Thirdly, we need to explain the apparent success of the policy of deregulation of the 1994-coalition. If we apply the three groups of theories of regulation, the following picture emerges.

According to the public interest theory, we would have expected the policy of deregulation to be successful in the eighties. The report of the committee on deregulation made abundantly clear that in many areas of policy, regulation appears to be ineffective, costly in financial terms both to the government and to the public and that over-regulation indeed threatens to undermine the very foundations of the constitutional state. Yet the policy of deregulation only seems to be effective in the nineties. How can we explain this?

The problem with the public interest theory is both the timing and the subjects of deregulation. Firstly, we would have expected deregulation to occur in the eighties. Secondly, the reduction and simplification of regulation mainly involves the markets for goods and services and not the labour market. We would have expected a successful policy of deregulation in this market as well. In this latter market many rules have been found to obstruct the processes needed for efficient market operation.

According to the Chicago theory, changes in the cost of information and organization in the nineties might have made it politically more rewarding to pursue a policy of deregulation. However, the gainers of a policy of deregulation did not constitute the government at the beginning of the nineties. Contrary to the eighties, when Liberals and Christian-Democrats constituted the government, in 1990-1994 the government coalition consisted of the parties of the Social-
Democrats and Christian-Democrats. The Social Democrat party is heavily supported by organized labour, which has much to lose by deregulation and the Christian Democrat Party is traditionally supported by the smaller businesses that generally profit from the establishment laws. Another possibility is that declining profits and income in the regulated sectors of the economy have made it impossible to uphold the regulation. However, according to the available statistics, the development of profits and income does not differ substantially between smaller businesses and the larger corporations.

Finally, we may seek an explanation for the failures and successes of deregulation in the political process of policy making. According to the 'behavioural' theories, either the supply of free information about the wealth enhancing effects of deregulation may have led to political support or deregulation has successfully been used as an issue in the elections. However, deregulation has not been an issue in the elections preceding the government coalitions of the eighties and nineties. Because of the complexities it is an unlikely subject to provide free information on to win the elections. Of course, in the Netherlands as in the rest of Europe, there has been a growing awareness of the limited possibilities of efficient and effective government policy. Market oriented political parties could win the elections and successfully pursue a policy of deregulation. However, if deregulation was implicitly an issue in the elections we would sooner have expected deregulation to be successful in the eighties when Liberals and Christian-Democrats constituted the government, rather than at the beginning of the nineties, when there was a coalition of Christian-Democrats and Social-Democrats. Furthermore, if there is uncertainty regarding the distribution of gains and losses of deregulation even a Pareto-efficient deregulatory policy might fail to receive adequate political support\textsuperscript{12}. This could explain why deregulation failed in the eighties, but it cannot explain why it was a comparative success in the nineties.

The 'structuralist' theories emphasize external factors such as changes in the balance of power of interest groups and changes in public opinion, that upset a structure-induced equilibrium of regulation and legislation. Indeed, there are

\textsuperscript{12} Fernandez and Rodrik (1991).
two such external factors that might explain the successes of deregulation in the Netherlands. Those factors are political entrepreneurship and a change in public opinion with an accompanying change in the balance of power of the political parties. In the Netherlands, the chairman of the committee on deregulation was in 1990 appointed secretary-general of the Department of Economic Affairs. This is the Department that prepares and implements competition policy and establishment regulations, precisely the areas that have been vigorously re-regulated or deregulated at the beginning of the nineties. The secretary-general is also a professor of European and Economic Law at the Erasmus University. As such he has published several articles on the policy of deregulation: what went wrong in the eighties and how should it be organized in the future to be successful? The Departments of Economic Affairs and Finance have, in cooperation with the Erasmus University set up a foundation (OCFEB) that finances research in the efficiency and effectiveness of financial and economic government policy. Many articles on regulation and deregulation have since then been published in journals and newspapers. This is one of the reasons why deregulation became once again a political priority. Finally, in preparing the dossiers for the formation of the cabinet in 1994, the secretary-general was in the position to influence the political agenda and the policy of deregulation.

The second factor upsetting the structure-induced equilibrium of regulation and legislation has been a change in public opinion and the accompanying change in the balance of power of the political parties. The elections of 1994 were won by the market oriented liberal parties who, together with the Social-Democrats formed the government coalition. For the first time since 1918 the Christian-Democrats do not take part in the government coalition. The Christian-Democrats have always been at the centre of the political spectrum and interest groups have naturally sought alliances with politicians of this party and with the bureaucracy. Before 1994, bureaucrats prevented the making of a policy of deregulation which was against their interests and politicians did not find it politically effective to press for decision making on deregulation. After 1994, a vacuum existed between interest groups and the ruling political parties which made it possible to pursue a policy of deregulation against the special interests. Furthermore, the institutional structure of the policy of deregulation was such that obstructions from the bureaucrats were practically ruled out. Firstly, a ministerial committee responsible is responsible for the preparation of
deregulation and not the minister of the department whose policy it concerns. Secondly, the working groups who analyse the areas of legislation and regulation include external experts and an independent chairman. Such a structure minimizes the available policy discretion for the bureaucrats whom it most concerns. Finally, implementation of a policy of deregulation is now more likely to succeed. Political decision making could be effectively supported by the preferences of the general public without possible distorting influences of special interest groups.

2.5 Conclusions

In the Netherlands the policy of deregulation has been on the political agenda since 1982. The policy of deregulation was not successful in the eighties but has been comparatively effective in the nineties. Of the three theories of regulation, the public interest theory, the Chicago theory and the public choice theory of regulation, the public choice theory seems to have the greatest explanatory power of deregulation in the Netherlands. The theory identifies the variables that may upset the structure-induced equilibrium of regulation and legislation. In the Netherlands two such variables can be identified: public opinion and political entrepreneurship. The first variable changed the balance of power of the political parties. The elections of 1994 were won by the liberal parties and for the first time since 1918 the Christian Democrat Party does not take part in the government coalition. Consequently, the interest groups find it difficult to stop the process of deregulation. The second variable, political entrepreneurship, was of crucial importance in the re-regulation of competition policy and the deregulation of the establishment regulation. The former chairman of the committee on deregulation was in 1990 appointed secretary-general of the Department of Economic Affairs, the department that formulates the proposals on the competition and establishment policies. As the head of the Department he could effectively control the bureaucracy and further the policy of deregulation which he deemed necessary.
CHAPTER 3    AN ECONOMIC ANALYSIS OF THE (SELF-)REGULATION OF PHYSICIANS IN THE NETHERLANDS

3.1  Introduction

In the context of the 1992 programme, the European Economic Community is aiming at a free flow of professional services. Regulation may hinder this freedom. This paper investigates the extent of (self-)regulation of the market for medical services. Regulations are analysed in terms of the characteristics derived from the neoclassical market model: do the rules restrict entry, competition and suchlike. This approach implicitly suggests that the relevant choice is between the ideal state of the neoclassical model, on the one hand, and an imperfect institutional arrangement on the other. This potential bias is counterbalanced by the analysis of the logic behind the rules. Are the regulations a result of market imperfections or market failures created by the special features of the services provided in question? And if they are not, can they perhaps be better explained by the rent-seeking activities of interest groups?\(^1\)

The first section of my paper defines the relevant medical professions and provides detailed information on the extent of (self-)regulation. Since proposals for radical reform are being discussed, which are to be implemented by the beginning of 1995, the final paragraph in this section describes the main features of these proposals. The second part of this paper analyses the logic behind the existing rules and investigates some of the market outcomes. Finally some conclusions will be drawn regarding the appropriateness of the rules in relation to their motives.

\(^1\) Explanations of regulations from either the ‘public interest’ or the ‘private interest’ point of view seem to be based on two contradictory concepts of the state. Rent seeking may be defined as attempts by individuals to increase their personal wealth while at the same time making a negative contribution to the net wealth of their community.
Chapter 3

3.2 The (self-)regulation of the medical professions

3.2.1 Definition of the medical professions

In the literature on economic activities, several criteria are deemed appropriate to distinguish professions from occupations, though none of them seems decisive\(^2\). This ambiguity in distinction also prevails between the medical professions and the medical occupations. Since this paper is concerned with the regulation of the medical professions we shall follow the distinctions made in those regulations pragmatically.

The medical professions consist of those who perform medical acts\(^3\). The law of 1865 (still operative) only distinguishes between physicians and obstetricians (midwives)\(^4\). Soon after this law several other medical professions developed in addition to the physician (such as the dentist for instance), which led to the introduction of new regulations. Up until the present day it is only the physician who is legally entitled to perform all medical acts.

Also within the profession of physician, specializations have developed, although none of them have been regulated in (supplementary) statutory laws. Three broad groups can be distinguished: general practitioners, medical specialists and social physicians (such as school doctors). Within the limited space of this paper it has been necessary to make a choice. This paper concentrates on the (self-)regulation of the profession of medical specialist. The first reason is that the medical specialist is the pivot in the second (polyclinical) and third (clinical) echelons in the Dutch health care system. Those echelons have undergone rapid growth in the last 15 years. The second reason is the limited success of a ten-year government policy to curtail the incomes of medical specialists.

\(^3\) Medical acts are defined by the law as the providing of curable, surgical and obstetrical services as a profession. It consist of the investigation of (a part of) the human body that does not work the way it should and of the recommendation of means to overcome the infirmity.
\(^4\) Wet regelende de uitoefening van de geneeskunst (1865).
Finally the profession of medical specialist as a whole seems to meet the characteristics of a rather successful pressure group.

3.2.2 Regulation and self-regulation of the profession of medical specialist

Market results arise from the combination of market conduct and market structure. Private interest theory of regulations suggest the determining of the market structure (in part) by successful pressure group activities. The basic elements of the market structure can be divided into three large groups:

- indirect entry barriers (restrictions on entrance to education and training etc.);
- direct entry barriers (such as certification, licensing, or outright physical planning etc.);
- elements determining inter and intra-professional competition (such as elasticity of demand, price and quality regulations, countervailing power, forms of co-operation and concerted action, professional codes of conduct etc.)

Barriers to entry and limitations on competition are generally considered to have unfavourable results in terms of technical, allocative and dynamic efficiency and income distribution. In the next three subparagraphs I will give a description of the extent of (self-) regulation concerning the aforementioned groups of market structure elements.

3.2.2.1 Indirect barriers to entry: education and training

Before entering specialist education one has to pass the examination in medical science. Since 1972 there has been a numerus fixus for the study of medicine. The numerus fixus was introduced by the Ministry of Education, at first, for
reasons of cost, but later for reasons concerning the labour market\textsuperscript{5}. The university study of medicine takes 6 years, including one year hospital practice (other sciences take about 4 years of study). Having successfully passed the examination in medical science and taken the oath, one is legally entitled to perform all medical acts, although of course, in reality, one is not fully qualified to do so. Each year 1,485 students (at the start of the eighties: 1,800) are allowed to enter the study of medicine, about 80\% (1,190) of whom pass the examination.

Unlike in other European countries, the education, recognition and registration of medical specialists is completely organised and performed by the profession itself (KNMG)\textsuperscript{6}. To perform this task there are three committees: a Central Committee (CC), a Specialist Registration Committee that keeps the medical register, and a Committee of Appeal. The CC investigates, among other things, the domain of new specializations, the possibility of making a living from a full time practice and the willingness of other specialists to leave the medical acts concerned to the new medical specialists. The composition of the CC is on the basis of parity: representatives of the eight university faculties and of the professional associations of medical specialists. The Minister of Education has the right to veto decisions made by the CC. To be registered as a medical specialist, one does not have to be a member of the professional association of physicians. The registration lapses after practising as a general practitioner and after five years of non-activity. At the moment there are 29 specializations which take 4 to 6 years of study and hospital practice. The candidate specialist has to find a recognized medical specialist willing to educate and train him. After completion of the education there is no examination: the candidate has to request registration as a medical specialist. In 1991 there were approximately 2,300 candidate medical specialists\textsuperscript{7}. This number of candidates is (strategically) determined by the profession itself on the basis of indicative manpower planning by the government, published in the Nota Beroepskrachten planning\textsuperscript{8}. Each year about 470

\textsuperscript{5} Ministerie van Welzijn, Volksgezondheid en Cultuur (1990) p. 15.
\textsuperscript{6} Koninklijke Nederlandse Maatschappij tot Bevordering der Geneeskunst (KNMG) (1981), art. 1001-1037.
\textsuperscript{8} Supra n. 5.
students are allowed to specialize, about 95% of whom round off the specialization⁹.

3.2.2.2 Direct barriers of entry resulting from (self-)regulation

The protection of the title and of the profession and an extensive definition of medical acts protects the profession against competition from other occupations or professions. Competition within the profession is potentially restricted by limitations on entry set by the profession itself, health insurance institutions and the government.

a. Protection of the title and of the profession

The title of “physician” (or doctor) is legally reserved for those who have taken the oath and passed the examination in the study of medicine¹⁰. Only the physician is entitled to perform (all) medical acts¹¹. By registering as a medical specialist the physician voluntarily agrees to restrict his domain to certain specialized medical acts. Stepping across one’s domain may even lead to a conviction, based on statutory disciplinary law. Specializations, on the other hand, imply the creation of sub-markets with monopoly rights (actually, not legally). Health insurance institutions and hospitals only sign their contracts with registered specialists. Before setting up as a specialist one has to have one’s certificate inspected by the mayor and by the Inspection of the Ministry of Health¹². This way the authorities know who is practising medicine legally.

---

⁹ Ibid, p. 41, 43.
¹⁰ Wet van 25 december 1878, Stb. 222, houdende regeling der voorwaarden tot de verkrijging der bevoegdheid van arts, tandarts, apotheker, verloskundige en apothekersbediende, art. 1 en 22 (eendaflegging).
¹¹ Supra n. 4, art. 1.
¹² Ibid, art. 4.
Chapter 3

b. Limitations to entry by the profession

The association of physicians (KNMG) has formulated ethical rules and professional codes of conduct limiting intra-professional competition. These rules are guidelines and have no validity on non-members of the association. Membership of the association is not compulsory. Private disciplinary sanctions on violating the rules are, for instance: the imposition of a fine of, at most, one thousand guilders and the striking off from the membership list\textsuperscript{13}. Also the association of medical specialists (Landelijke Specialisten Vereniging: LSV) has formulated specific codes of conduct for medical specialists.

In principle, there are two ways of setting up in practice: with or without associating with fellow specialists\textsuperscript{14}. Setting up a practice (without association) is not allowed within a two-year period of standing in for a colleague in that same area. It is also forbidden to set up practice after acquiring knowledge of the value of a practice in negotiations with local specialists for a take-over or an association. Finally a specialist is not allowed to set up a practice in the area of his tutor. One of the rules declares the undesirability of 24-hour personal care by a physician for his patients. The co-operation with colleagues for stand-in during the night, weekend and holidays is thus almost always necessary. Before starting a practice a specialist therefore has to inform the local organization of specialists of his plans\textsuperscript{15}. Participating in the association of fellow specialists requires the payment of goodwill, according to the rules, amounting to about one year’s income\textsuperscript{16}.

\textsuperscript{13} KNMG (1981), articles 600-682.
\textsuperscript{14} A small minority of medical specialists are working as remunerated employees in hospitals.
\textsuperscript{15} KNMG (1984) articles 61-69.
\textsuperscript{16} LSV (1990).
c. Limitations to entry by the health insurance institutions.

Medical specialists receive their fees either from patients (in the case of private insurance) or from public insurance institutions\(^\text{17}\). Over 60% of the Dutch population participates in public insurance\(^\text{18}\); a salary of under Hfl. 54,500 (1991) renders participation obligatory. The public insurance institutions are obliged to contract each registered specialist who wishes to, except where there are ‘serious objections’\(^\text{19}\). Examples of such objections are the lack of stand-in arrangements or clinical facilities. The institutions can also ask the Minister of Health to be relieved from the obligation to contract. According to the law\(^\text{20}\) the minister can do so if the public insurance institutions have already made sufficient contracts with medical specialists to fulfil their duties\(^\text{21}\). Basic elements of the contract have been written down in law and standard form contracts have been agreed upon by the organizations of medical specialists (LSV) and the public insurance institutions (Vereniging Nederlandse Ziekenfonds: VNZ).

d. Limitations to entry by the government

The practising of almost every medical specialization requires a hospital. It has diagnostic facilities not available elsewhere (X-ray examination, laboratory), also diagnosis and therapy depend on the presence of hospital beds and an increasing number of patients need multidisciplinary treatment. The relationship between the hospital and the medical specialist is governed by the admission

\(^{17}\) According to the public insurance law insures have a right to medical services in kind. Private insurance restitutes the amount of money payed by the insurers for the rendered medical services.

\(^{18}\) Supra n. 7, p. 37; it concerns 9,117 million people.

\(^{19}\) Ziekenfondswet, art. 47 lid 1. The obligation to contract formally aims at the guaranteeing patients the freedom of choice of physicians. Materially, the organisation of physicians compelled them not to contract with insurance institutions who did not effect this obligation in their statutes. See, Leenen en Roscam Abbing (1986), pp. 112, 134.

\(^{20}\) Ibid, art. 47 lid 3.

\(^{21}\) So far the public insurance institutions have not asked to be relieved of the obligation to contract with medical specialists. The minister has granted the exemption for contracting with dentists and physiotherapists.
Chapter 3

contract and the medical staff regulations. For both, standard form contracts have been agreed upon in negotiations between the organization of medical specialists (LSV) and the organization of hospitals (Nationale Ziekenhuis Raad: NZR)\textsuperscript{22}.

The number of hospitals is legally determined by a planning system for the 27 health regions\textsuperscript{23}. Preparations are made by the regional authorities on the basis of ministerial guidelines. Those guidelines contain specific norms such as, for example, 3.4 beds per thousand persons. Every hospital has to have primary and supporting specializations to offer help 24 hours a day. The required number of specialists is measured in what are called function units: the annual production of a specialist on a full-time basis. The function units differ for the various specializations and are determined on the basis of the size and age profile of the regional population. For the nine top specialist functions (incl. heart surgery and kidney transplants) there is a licence system\textsuperscript{24}. Hospitals are forbidden to perform those operations or purchase equipment unless they are licensed to do so. A licence usually permits only a specified number of operations.

When a hospital fits into the planning scheme, the Minister of Health grants approval to the hospital. This so-called “recognition” contains function units, beds, personnel and equipment an, since 1982, even numbers of specialists\textsuperscript{25}. If a hospital wants to increase the number of specialists, it needs permission from the minister. The financing of a hospital from public insurance funds requires ministerial recognition. This ministerial recognition forms the basis for the maximum expenditure that hospitals are allowed to make. Since the end of 1989 the Minister of Health has the legal authority to close down a hospital, to reduce the number of beds and function units of specialists and to end the practising of a specialism within a hospital\textsuperscript{26}.

\textsuperscript{22} Latest version, see Vijfpartijen Akkoord (1991).
\textsuperscript{23} Wet Ziekenhuisvoorzieningen, 1971, 1979.
\textsuperscript{24} Ibid, art. 18.
\textsuperscript{25} Staatscourant, 1982, 132.
\textsuperscript{26} Tijdelijke wet capaciteitsreductie ziekenhuizen, TK 21200, 28 juni 1989.
Finally by virtue of the health care facilities law\textsuperscript{27} the government has the authority to forbid the practising of a profession without a licence\textsuperscript{28}. The establishing of additional sectors or their expansion can be prohibited in certain parts of the country due to current oversupply, and in order to stimulate the practice elsewhere. The maximum and minimum numbers of medical specialists (or other professions) may be determined through the granting of licences as well as by the maximum number of patients\textsuperscript{29}.

3.2.2.3 Regulation and self-regulation of competition

Competition manifests itself in price and quality adjustments to (changes in) demand and supply. Competition may be restricted through regulations on price and quality. Competition may also be hindered by direct forms of collusion or co-operation.

a. Price regulations.

The determination of price in the health care sector is regulated by law\textsuperscript{30}. In principle, prices in the health care sector result from negotiations between the suppliers of the care and the insurers. There is a public institution (COTG) that has the authority to approve the prices agreed upon or to determine the prices themselves\textsuperscript{31}. No new or higher prices in the private or public sector may be charged without its approval. The COTG consists of 18 members: 6 appointed by the government and 6 members each, appointed on the recommendation of employers and employees, the health insurers and the suppliers of health care. For its work the COTG has six advisory ‘chambers’ among which chamber 5 for physicians, dentists and pharmacists. The determination of prices by the COTG rests upon guidelines that need the approval of the Minister of Health. The min-

\begin{itemize}
\item\textsuperscript{27} Wet voorzieningen gezondheidszorg (1982).
\item\textsuperscript{28} Ibid, art. 18.
\item\textsuperscript{29} So far a system of licence has only been introduced for general practitioners.
\item\textsuperscript{30} Wet tarieven gezondheidszorg (1980).
\item\textsuperscript{31} COTG: Centraal orgaan tarieven gezondheidszorg.
\end{itemize}
ister can also give specific instructions concerning the content of the guidelines. Fees for medical specialists contain two components: income and costs. The COTG is not allowed to formulate and execute an incomes policy. There is a special law for the determination of incomes. According to this law the minister can determine so-called acceptable income levels after consultation with the profession. The acceptable income levels of specialists result from the comparison of a normative income and an actual income. This special law forms the basis of specific ministerial instructions concerning the income part in the tariffs. Specific instructions concerning the cost part have their basis in yet another law.

If prices were not centrally determined, price rigidity would also originate from the code of conduct from the profession. The rules stipulate physicians to maintain the fees agreed upon or accepted by the professional organization.

b. Regulations on quality.

Quality is a difficult concept. It might be defined as the extent to which a complex of attributes of a product meet the requirements necessary for its use. Quality judgement may concern the structure, the procedures and the outcomes of the medical services rendered. Many of the aforementioned forms of self-regulation are intended to improve quality. Generally speaking one might say that an improvement in quality is furthered mostly by regulating structure, although procedures are becoming increasingly important (protocols for diagnostics and therapy, testing among colleagues, promotion of consensus). The judging of quality by outcomes has not been regulated in statutory law nor in codes of conduct by the profession.

---

32 Originally, before 1980, the fees for specialists were kept outside price regulations for hospitals on the grounds that: there are no accurate norms available for comparison with other professions, it would make no sense without an obligation for specialists to contract with the insurance institutions and finally price control by the government would not be desirable or feasible. See Leenen, Roscam Abbing (1986) p. 163.


34 Prijzenwet (1961).

35 KNMG (1984), art. 35.
c. Regulations on co-operation, partnership and competition.

Competition among physicians is opposed as much as possible by the organization of physicians (KNMG). The rules of conduct stipulate that physicians should be reserved in taking over a patient from a colleague in the same working area. It is not permitted to accept a patient within six months after helping this patient while standing in for a colleague36. The international code of medical ethics simply states that a physician shall not entice patients from his colleagues37. The professional ethics also contain a general prohibition on the advertising of professional skills or specialized methods of treatment38. Advertising should be limited to three times within the month of setting up a practice. According to the international code of medical ethics, self-advertising is deemed to be unethical conduct. The KNMG-code of conduct emphasizes the professional autonomy of physicians: co-operation should be in the form of an association, hierarchical structures among doctors is undesirable39. The association contract requires the approval of the professional organization (LSV)40. When after the novicite period the medical specialist decides not to join the association, he is not allowed to practice within a 10 to 20 km area around the hospital where the association is to be found41. An association of different specialists is only possible under specified conditions42. After joining the association, termination of the agreement is not possible within a three to five year period. The association functions on the basis of financial equity; additional earnings (expert opinion for life insurance companies etc) also have to be split up and divided into equal parts43.

36 Ibid, art. 52.
38 KNMG (1984) articles 82-84.
39 Ibid, art. 73.
41 LSV (1979), art. 2.9 and p. 19.
42 The special rules aim at preventing doubling in the expense accounts and unnecessary referring to colleagues in other specializations.
43 LSV, supra n. 40, art. 3 and p. 10.
Chapter 3

An antitrust law came into existence in 1958. Originally this law was not intended for these professions, but in 1987 the law was extended to include them. Regulations restricting competition, such as the codes of conduct, have to be reported and registered. When they are not in the general interest they can be declared invalid in law. Ministerial directives (on prices, supply, admission of a specialist or a stand-in arrangement etc.) can be issued when suppliers occupy a power position on the market.

3.2.2.4 Reform proposals

Several changes are currently being proposed by the government and discussed in parliament. In the near future the insurance institutions will have genuine freedom of contract. The difference between private and public insurance will disappear. The regional monopoly position of the public insurance institutions will be abolished. The insured rights to medical services will be written in functional terms instead of services provided by registered specialists in recognized establishments. Instead of a system of protection of the title and of the profession there will solely be a system of protection of titles. Only specific medical acts will be reserved for physicians or medical specialists. The titles of specialists will be protected in statutory law. Prices (and qualities) of medical services will be determined by negotiation, although there will be upper ceilings. Horizontal price agreements will be prohibited. The law on acceptable incomes of these professions will be revoked. The planning system for hospitals will be deregulated and decentralized; the costs of medical specialists will be made part of the hospital budget.

3.3 The economic logic of (self-)regulation of the profession of medical specialists

This part of the paper investigates the logic of regulations. Paragraph 3.3.1 investigates whether the regulations can be seen as responses to market failures

---

44 Wet economische mededinging (1958).
or to market imperfections. The next section considers the interest group theory of regulations: can the rules be explained from pressure group activities.

3.3.1 The logic of regulations: the public interest theory

Several motives seem to underlie the extent of (self-)regulation in the market for specialized medical services. Historically a subdivision is possible between primary and secondary motives. There are two primary motives for (self-)regulation: the provision of accessible health care and the guaranteeing of the quality of that health care. Resulting regulations and market outcomes gave rise to two secondary motives: the macro-economic control of the health care sector and an acceptable income distribution. I shall discuss these motives and the regulations based upon them respectively.

The regulations based upon the two primary motives significantly determined the market structure for specialized medical services. Firstly, they resulted in an unrestricted (induced) demand for medical services. Secondly, on the supply side, self-regulation of the profession created a monopoly position on price, quality and quantity.

3.3.1.1 The provision of accessible health care

Generally speaking, consumers are uncertain as to when they will need health care. Being, on average, averse to risk, they will want to insure this risk. Private insurers will offer insurance but they will try to identify the (un)favourable risks. The results of market forces are deemed unacceptable in two ways: the access to health care is dependent on income and better health care is available for those with higher incomes. According to the constitution, the government is obliged to promote public health (art. 22). Access to health care has been facilitated by compulsory public health insurance for employees with a salary of under Hfl. 54,500 (1991) and by insurance premiums as a percentage of income. In
addition to distributional considerations, there were two other reasons for compulsory public insurance. Firstly, to prevent a ‘free ride’ on tax payers’ money. If health care were required, non-insurance or underinsurance would mean an appeal to the Bijstandswet, which supplies the necessary costs of living. Secondly, external effects such as the employer’s interest for timely and adequate medical treatment require compulsory insurance.

The financial accessibility of health care presupposes the geographical and functional availability of health care. This is another reason for government intervention in the market. Health care facilities are so-called mixed goods: they have both public good and private good characteristics. Insofar as health care facilities have an availability utility (as distinguished from user utility) they are public goods and require collective action.

### 3.3.1.2 The guaranteeing of the quality of health care

The second motive for (self-)regulation has been stated in economic terms by Akerlof. Informational asymmetry can lead to certain types of market failure. Patients have difficulty in distinguishing the relative qualities of medical specialists. Since all specialists will be viewed as perfect substitutes, patients are willing to pay a price reflecting the average quality. The specialists themselves know their own abilities and alternative opportunities for employment. Future medical specialists with above-average opportunities may not be willing to enter or stay in the market, thereby lowering the average quality (adverse selection). The patients must know that all the medical services they can buy must be of below average quality, so they would be paying too much for them and would not buy. The market might break down altogether, despite the fact that if information were perfect, many transactions would take place.

Given the imperfection of limited information, a government can correct this imperfection in several ways. In the Netherlands a system of self-regulation

---

46 Akerlof (1970); see also Leland (1979).
(price, quantity, quality) has been introduced for both principle and practical reasons. The principle reason is that self-regulation is a way of protection against the government\textsuperscript{47}. Practical reasons are the presupposed comparative advantages in:

- ‘state-of-the-art’ information and expertise;
- flexibility to altered circumstances;
- precision and differentiation in the choice of instruments;
- vindication of the promulgated norms and rules\textsuperscript{48}.

The combination of self-regulation and public health insurance led to a rapid growth of the health care sector\textsuperscript{49}. On the supply side, medical specialists (physicians) acting in accordance with professional standards, experienced no restrictions in offering patients the best of help. New technologies were rapidly introduced. On the demand side, serious moral hazards developed from the particular design of the health care insurance. Patients did not have to pay (directly) for medical services. The obligation to contract and the professional autonomy of medical specialists made it practically impossible for the insurance institutions to counterbalance their position. This situation induced two secondary motives for regulation: macro-economic control of the health care sector (3.3.1.3) and an acceptable income distribution (3.3.1.4).

\textsuperscript{47} Self-regulation as a method of decision making has a long standing tradition in the Netherlands. It has religious roots in the (protestant) principle of ‘sphere sovereignty’ and the (catholic) principle of ‘subsidiarity’.

\textsuperscript{48} Geelhoed (1987), p. 15.

\textsuperscript{49} Labour intensity, little or no possibility of productivity growth and uniform wages are another (almost inevitable) cause of the increasing share of health care sector, see Baumol (1967). The growth of wages in the health care sector in accordance with the average wage growth has been regulated in policy and in law (Wet arbeidsvoorwaarden gepremieëerde en gesubsidieerde sector; WAGGS).
Chapter 3

3.3.1.3 The macro-economic control of the health care sector

The development of the health care sector had detrimental effects on the economy. Rising insurance premiums created an increasing divergence between net and gross income which have negative incentive effects. Furthermore, public insurance premiums are being compensated for in wage negotiations. The resulting price/wage spiral has unfavourable effects on the competitiveness and growth of the national product. The economic scrapping of machinery accelerates, and technical progress will be of the labour saving type. Profits, investments and employment will decline. Macro-economic control implies, at the micro level, the control of costs and the pursuing of an efficient allocation of scarce resources. To this aim, planning laws (WZV, WVG) and price laws (WTG: COTG) were promulgated.

3.3.1.4 Acceptable incomes and a balanced income distribution

The price law in the health care sector (WTG, 1980) had two objectives. First to restructure the price system: until then two prices prevailed for the same medical act, one in the public insurance sector and one in the private insurance sector. The second objective was to control the costs of the health care sector. The law on the acceptable incomes of these professions (WIVB) was instrumental to the price law and had as principal objectives: the control of the level and distribution of the income of medical specialists. The control of incomes of these professions was also necessary to prevent increasing differences of income between the private and the public sectors and within the health care sector itself. A table, combining regulations and objectives, while differentiating between regulation and self-regulation, can be found in Appendix I.

---

50 Two different prices for one and the same treatment is considered unjust. This differentiation originated from the poor-relief, but it still exists. Generally speaking prices (fees) in the private sector are twice as high as in the public sector. Over 60% of the population participates in public insurance. Yet the costs of specialist medical help financed by public insurance is 55% (FOZ, 1992, bijlage 11, blz. 271). This can be explained by a larger consumption of medical services in the public sector.
3.3.2 The logic of regulations: the private interest theory

Whether the regulations can be considered to be the result of pressure from interest groups is a difficult question to answer. Firstly, I shall discuss the question as to whether the organization of medical specialists meets the characteristics of pressure groups as described in theory. Secondly, I shall investigate whether the system of (self-)regulations contains irregularities that might be ascribed to pressure group activities. In the third place I shall analyse market outcomes resulting from self-regulation.

3.3.2.1 Medical specialists as a pressure group

Individuals can maximize utility either through productive activities within the existing framework of regulations or by seeking favourable changes of those regulations. The second path requires transaction costs, overcoming the ‘free rider’ problem and asymmetrical information on the distributional effects of the change in regulation. Small groups, with a possibility to commit their members, are in a better position than individuals or large groups for minimizing on transaction costs and preventing ‘free riding’. Asymmetrical information usually prevails when groups organize around single issues. The national association of medical specialists (LSV) fits these characteristics perfectly. It is a small group (about 7,500 members), traditionally well organized (membership level of 85%) and including disciplinary law\footnote{LSV (1990), Jaarverslag 1988/1989, blz 17.}. The rules of conduct stipulate physicians should devote themselves to looking after the material interests of the profession\footnote{KNMG (1984), art. 54.}. Information on medical subjects as well as its real costs is asymmetrically divided and often not even known\footnote{Many conflicts between the government and the LSV concerned the information on incomes (and thus costs) of medical specialists. This information is necessary to apply the law on incomes (WIVB). When finally, after five years, an accountant report had been drawn up, the LVS, through a lawsuit, succeeded in preventing publication.}. An illustration of their power is the General Agreement they were able to make (in 1985) with the government,
setting aside parliament and statutory laws. Ironically, when it came to executive measures, the LSV succeeded in making it invalid in law through a lawsuit.

Recently their power seems to be diminishing. Chest surgeons and radiologists have left the LSV and alternative associations have developed: the Netherlands Specialisten Genootschap on the ‘left’, and the Nederlandse Specialisten Federatie on the ‘right’ and the Academische Specialisten Vereniging. Countervailing power has also developed. There are currently about 175 categorial and 100 general organizations for patients\(^5\). The Landelijk Patiënten / Consumenten Platform (founded in 1986) succeeded in getting an action by the LSV (‘Sunday services’) forbidden in a lawsuit.

### 3.3.2.2 Anomalous features in health care regulations

There seem to be at least three irregularities in the system of health care regulations which would indicate the possible outcomes of pressure. First of all, hospitals fall under a system of planning whereas the supply of medical specialists can only be influenced indirectly. Taking the planning system as given, allocative efficiency would seem to require coordinating hospital facilities and medical specialists. Under the assumption that supply creates its own demand\(^6\) the planning system could eventually break down. Waiting lists, creating pressures on the planning system, originate as a result of the discrepancy in the method of control. Secondly, medical specialists operate on a ‘for profit’ basis in a not-for-profit hospital. They can increase income by expanding their output while hospitals operate on a limited budget. Expanding output for a hospital means increasing costs. As a consequence the hospital has to economize on other departments\(^6\). Thirdly, allocative efficiency would seem to lead to a system of


\(^6\) The distorting effects on allocation are being institutionalized since medical specialists are going to take part in the management and control of hospitals.
manpower planning in accordance with the hospital planning. Yet the profession of medical specialists is the only profession in the system of health care that regulates the system of education and training itself.

3.3.2.3  Competition and market outcomes

Another indication of interest groups and rent-seeking might be found in the outcomes of the (self-)regulations: is workable competition, in fact, eliminated and is the income of the medical specialist above average?

a. Competition and countervailing power

Looking at the self-regulations concerning education, practising and associating, intraprofessional competition among medical specialists seems to be eliminated. This is due to the fact that co-operation among colleagues in hospitals is necessary. Whenever there is room for another specialist, the specialist first applies the association and then to the hospital. Many hospitals, in fact, require participation in the association. Competition might also be enhanced by the establishment of private clinics57. It is difficult for specialists to associate within a private clinic. The financing of medical services by the public insurance institutions requires, in principle, ministerial approval of the clinic. This approval will not be given since it would thwart the planning procedures. Besides, according to the law (WZV), clinics wanting to treat patients for more than one day automatically fall under the planning system. At this moment only about 30 private clinics have been established (for example plastic and eye surgery, orthopaedics etc.). Legally, interprofessional competition has been eliminated since physicians have a monopoly on performing medical acts. Materially, the so-called alternative medical science (acupuncture, homeopathy etc.) has expanded rapidly. The number of ‘doctors’ practising illegally is estimated at over one thousand; the number of illegal medical acts is estimated at about 7 million58. Proba-

57 A private clinic is competition for the hospital. Hospitals are found to react immediately on the establishment of a private clinic by extending the consulting hours into the evening.
Chapter 3

bly most of these acts compete with the work of a general practitioner rather than a medical specialist but further research should make this clear. Competitive pressures might also be expected from the countervailing power of patient’s organizations. I mentioned a striking example above but once again further research is necessary to draw general conclusions. Finally, countervailing power may be exerted by the organizations of public and private insurance institutions. Price negotiations take place before the COTG approves the agreements. In practice, however, these negotiations have not prevented the emergence of huge incomes for medical specialists. This result might be explained by two factors.

Firstly, the number of candidate specialists is limited to prevent a weakening of the position of the LSV. The second factor is the lack of incentives to curtail the cost of medical help. This is because the medical professions themselves are represented in the administration of the public insurance institutions. Also a rise in cost would be distributed among several millions of employees through a small increase in the premium. The leading principle in negotiations might thus be the minimization of conflict rather than cost. The same reasoning applies to the organization of private insurers. Competition among the insurers would not be distorted by a uniform price (irrespective of its level). The COTG would find it difficult to disapprove of the several thousands of prices agreed upon, first of all because of transaction costs, secondly because it is dependent on the information given by the contracting partners (information costs).

b. Quality

Uniform prices do not guarantee uniform quality. Research into the differences in the use of medical services, demonstrates large differences in ‘production’ of the medical specialist. For instance after correcting for demographic and socio-economic circumstances, specialists in one region diagnose up to twelve times as many heart and lung complaints as in other regions. Echo-cardiograms are
made fourteen times more often and wombs are being removed up to four times as much\(^\text{59}\).

c. Incomes

Under the prevailing system of self-regulation and given the system of administrative prices, medical specialists have several opportunities to increase their income\(^\text{60}\). They could try to perform more medical acts, for instance by working longer hours, by substituting to standardized routine acts or by delegating medical acts. The latter is possible either through the so-called ‘lengthened arm’ procedure: (male) nurses performing (parts of) medical acts under the supervision of the specialist or by increasing the number of trainees. In both cases specialists receive the fees for the medical acts performed, although, in the case of trainees, they have to pay a small salary. Increasing the number of trainees would ceteris paribus eventually undermine the market power of the specialists. To avoid this effect the LSV has created a new function: the so-called candidate specialist not in training (agnios) in addition to the candidate specialist in training (agios). Increasing income is also possible by switching to privately insured patients or to certain types of medical acts with a high tariff. Increasing income is also possible by introducing labour saving technology, since fixed prices are rigid. Finally, increasing net income is possible by using cheaper material and instruments.

The income of medical specialists can be defined as the received fees with costs deducted; given as a formula:

\[
\text{Price} \times \text{Volume} = \text{annual turnover} = \text{Income} + \text{Cost}
\]

There are considerable difficulties in calculating the incomes of medical specialists. For instance, what can be considered the costs of a practice: is a Volks-


\(^{60}\) The assumptions of maximizing income is a rather narrow one. In the literature, the assumption is that a physician maximizes his utility as a function of income, working conditions, leisure, ethical concern and prestige.
wagen sufficient or is a Bentley adequate means of transportation. We shall simply refrain from these and other difficulties and present a few illustrating figures on the financial position of medical specialists\textsuperscript{61}.

In 1983 the average income of the 7,100 free practising medical specialists was about Hfl. 257,000. With this income the medical specialists is part of the top 10\% of entrepreneurs (!) who earned an average income of Hfl. 177,400 in 1983 (average income of all entrepreneurs: Hfl. 52,000). Among medical specialists there are large differences. For instance 600 children’s doctors earn an average income of Hfl. 155,000 while 600 radiologists earn Hfl. 360,000. There are also considerable differences (even within specializations) depending on the juridical structure opted for by specialists. For instance the average income of a partnership in the form of a closed corporation is almost 50\% higher than that of the ordinary partnership.

The differences in income between different specializations seem to originate from two sources. First, generally speaking, there are two kinds of medical acts: the ones taking mostly labour time (so-called ‘non cutting’ branches) and the ones taking labour time and equipment (‘cutting’ branches). The latter have undergone rapid increases in productivity. There is no apposition of once determined tariffs apart from the compensation of rising prices and wages. Increasing production and given tariffs thus generate increasing income. The second source of differences in income is the difficulty in defining different medical acts when diagnosis and therapy consists mainly of labour time. Using equipment leads to more differentiation in medical acts, more expense accounts and thus more income.

\textsuperscript{61} For more details, see the statistical annex.
3.4 Concluding remarks

The complexity of the subject and the lack of necessary empirical knowledge prevent the drawing of conclusions in a more systematic and precise way. Having said that, the following comments are deemed appropriate.

1. There is no doubt that the organization of medical specialists fits the characteristics of effective pressure groups as described in theory. In the context of the present government policy their activities might more accurately be described as rent keeping than rent-seeking. The current average income of medical specialists seems excessively high, although there is a fair amount of variation in the distribution of income among different specializations.

2. Looking at the number of (self-)regulations, one might conclude that the market structure is decisively determined, not so much by the extent of self-regulation of medical specialists, but by the regulations of the demand side (insurance) of the market. Those regulations open up possibilities that would not otherwise have existed. Insurance institutions properly (de)regulated should be able to overcome the information problems and stimulate competition in the market for specialized medical services.

3. A precondition for stimulating competition is the abolition of several forms of (self) regulation. Specifically the obligation for insurance institutions to contract with medical specialists prevents workable competition. Another barrier to workable competition is the determination of the capacity for educating and training of candidate specialists by the profession itself. Thirdly, current planning regulations consolidate the position of medical specialists and hinder competition; for instance insurance institutions should be able to introduce alternatives such as the ‘preferred supplier organization’ or ‘health maintenance organizations’. Finally, freedom in the determination of prices and qualities should be established. After implementation of these alterations, traditional competition policy should be sufficient to guarantee workable competition.

4. Informational asymmetry is the main practical reason for a system of self-regulation. Self-regulation should guarantee quality and facilitate consumers in
making suitable choices. Apart from quality control and information by the insurers, there are other alternatives to self-regulation, such as accreditation, certification, legal obligation to disclose certain information etc. In the economic literature there seems to be no unanimity as to the efficiency of self-regulation.

5. As a consequence of the plurality of objectives of regulations it is difficult to analyse the appropriateness of separate regulations as responses to specific market failures. A change in regulation would not just alter specific market outcomes in terms of one objective but possibly also in terms of another. There are trade-offs to be expected in the pursuing of quality and cost, income distribution and quality, cost and accessibility etc. A judgement of appropriateness would require knowledge of the marginal rate of substitution between the objectives. This knowledge is not available.

62 For instance: ‘Extremely strong support for the argument that the granting of monopolistic powers to self-regulating professions is likely to be welfare reducing…’ (Shaked and Sutton (1981), p. 223); ‘…in at least one dimension the restriction on entry may prove to be efficient’ (Barzel, 1982, p. 38).
Appendix 1: Regulations and Objectives

<table>
<thead>
<tr>
<th>OBJECTIVES</th>
<th>QUALITY</th>
<th>ACCES</th>
<th>COST</th>
<th>DISTRIBUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGULATION (public)</td>
<td>WWO (education)</td>
<td>Ziekenfondswet (obligation to contract)</td>
<td>WWO (numerus fixus)</td>
<td>WTG</td>
</tr>
<tr>
<td></td>
<td>WUG (definition protection of titel/profession)</td>
<td>Wet ziekenhuisvoorzieningen (function units)</td>
<td>Ziekenfondswet (relief from obligation to contract)</td>
<td>WIVB (normative income; acceptable income)</td>
</tr>
<tr>
<td></td>
<td>Wet 1878 (exams; oath)</td>
<td>Wet Voorzieningen Gezondheidszorg (licence)</td>
<td>Wet ziekenhuisvoorzieningen (art.18: top clinical functions)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ziekenfondswet (serious objections, art. 47.1)</td>
<td></td>
<td>Wet voorzieningen gezondheidszorg (licence)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ziekenfondswet (ministerial approval)</td>
<td></td>
<td>Prijzenwet (obligations on administration and calculations)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KNMG (ministerial veto)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wet economische mededeling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Registration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ministerial directions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF-REGULATION (private)</td>
<td>KNMG (education and register)</td>
<td>LSV (association of different specialisms)</td>
<td>LSV (associations of different specialisms)</td>
<td>LSV</td>
</tr>
<tr>
<td></td>
<td>KNMG (conditions for establishment and stand-in arrangements)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSV (standard form contracts with hospitals and insurers)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KNMG code of conduct (advertisement, competition associating)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSV (association)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3

Appendix 2: Statistics on Medical Specialists

Table of contents (the numbers refer to the tables):


10. Numbers of candidate medical specialists each year allowed to enter education and training (Ministerie van Welzijn, Volksgezondheid en Cultuur, 1990, p. 4).


12. Incomes of medical specialists based on 1977-statistics (Goudriaan, p. 113).

13. Incomes of medical specialists based on 1977-statistics (Goudriaan, p. 113).
An Economic Analysis of the (Self-)Regulation of Physicians in the Netherlands

Table 1: Expenditure on health care (x mld gld) as a percentage of GNP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EHC</td>
<td>20,0</td>
<td>32,9</td>
<td>40,5</td>
<td>41,9</td>
<td>42,9</td>
<td>44,1</td>
<td>45,6</td>
<td>48,4</td>
<td>51,4</td>
<td>52,6</td>
<td>58,1</td>
</tr>
<tr>
<td>GNP</td>
<td>219,9</td>
<td>336,1</td>
<td>418,9</td>
<td>427,6</td>
<td>429,4</td>
<td>447,4</td>
<td>476,0</td>
<td>509,0</td>
<td>536,6</td>
<td>560,8</td>
<td>630,9</td>
</tr>
<tr>
<td>%</td>
<td>9,1</td>
<td>9,8</td>
<td>9,7</td>
<td>9,8</td>
<td>10,0</td>
<td>9,9</td>
<td>9,6</td>
<td>9,5</td>
<td>9,6</td>
<td>9,4</td>
<td>9,2</td>
</tr>
</tbody>
</table>

Table 2: Financing of the expenditure on medical specialists (mln gld)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Insurance</td>
<td>964</td>
<td>976</td>
<td>879</td>
<td>841</td>
<td>1017</td>
<td>993</td>
<td>1.119</td>
</tr>
<tr>
<td>Private Contributions</td>
<td>994</td>
<td>46</td>
<td>168</td>
<td>233</td>
<td>63</td>
<td>62</td>
<td>81</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>957</td>
<td>975</td>
<td>949</td>
<td>1029</td>
<td>1005</td>
<td>790</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>62</td>
<td>74</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum Total</td>
<td>1,958</td>
<td>1,979</td>
<td>2,022</td>
<td>2,085</td>
<td>2,183</td>
<td>2,134</td>
<td>2,063</td>
</tr>
</tbody>
</table>

Table 3: Number of insured persons (x 1000)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>14,572</td>
<td>14,665</td>
<td>14,760</td>
<td>14,849</td>
<td>14,951</td>
</tr>
<tr>
<td>Public Insurance</td>
<td>9,083</td>
<td>8,973</td>
<td>9,011</td>
<td>9,134</td>
<td>9,177</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>5,550</td>
<td>5,604</td>
<td>5,518</td>
<td>5,556</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Patients consulting medical specialists and admittance to a hospital

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Consultations p.p.</td>
<td>1.6</td>
<td>1.6</td>
<td>1.8</td>
<td>1.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Inhabitants (x 1000)</td>
<td>14,491</td>
<td>14,572</td>
<td>14,665</td>
<td>14,760</td>
<td>14,849</td>
</tr>
<tr>
<td>Nos of Consults (mln)</td>
<td>23.2</td>
<td>23.3</td>
<td>26.4</td>
<td>23.6</td>
<td>28.2</td>
</tr>
<tr>
<td>Admittance to a Hospital (x mln)</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 5: Percentage of insured consulting a medical specialist

<table>
<thead>
<tr>
<th></th>
<th>1981</th>
<th>1989</th>
<th>1990</th>
<th>Annual Rate of growth</th>
<th>Rate of Growth in 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Insurance</td>
<td>35.2</td>
<td>38.9</td>
<td>42.0</td>
<td>1.3</td>
<td>8.0</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>37.1</td>
<td>37.4</td>
<td>39.7</td>
<td>0.1</td>
<td>6.1</td>
</tr>
<tr>
<td>Total (average)</td>
<td>35.7</td>
<td>38.2</td>
<td>41.0</td>
<td>0.8</td>
<td>7.3</td>
</tr>
</tbody>
</table>
Chapter 3

Table 6: Number of medical specialists practising in hospitals (in function units)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General Hospitals</td>
<td>5.680</td>
<td>5.730</td>
<td>5.766</td>
<td>5.784</td>
<td>104</td>
</tr>
<tr>
<td>Categorial Hospitals</td>
<td>474</td>
<td>491</td>
<td>428</td>
<td>394</td>
<td>-80</td>
</tr>
<tr>
<td>Academic Hospitals</td>
<td>2.075</td>
<td>2.075</td>
<td>2.075</td>
<td>2.075</td>
<td>0</td>
</tr>
<tr>
<td>Sum Total</td>
<td>8.229</td>
<td>8.296</td>
<td>8.269</td>
<td>8.253</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 7: Position of medical specialists

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered Medical Specialists</td>
<td>10796</td>
<td>11206</td>
<td>11612</td>
<td>11957</td>
<td>12265</td>
<td>12689</td>
</tr>
<tr>
<td>- Below the age of 65</td>
<td>9955</td>
<td>10247</td>
<td>10534</td>
<td>10764</td>
<td>10973</td>
<td>11088</td>
</tr>
<tr>
<td>Of whom:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Contracted with Public Insurance Companies</td>
<td>6979</td>
<td>7132</td>
<td>7117</td>
<td>7163</td>
<td>7293</td>
<td>7500</td>
</tr>
<tr>
<td>- Practising in General Hospitals</td>
<td>5723</td>
<td>5680</td>
<td>5730</td>
<td>5766</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>- Practising in Academic Hospitals</td>
<td>1789</td>
<td>1840</td>
<td>1971</td>
<td>1995</td>
<td>2085</td>
<td>2132</td>
</tr>
<tr>
<td>- Practising elsewhere</td>
<td>642</td>
<td>633</td>
<td>611</td>
<td>554</td>
<td>549</td>
<td>565</td>
</tr>
</tbody>
</table>

Table 8: Number of newly registered medical specialists and number of medical trainers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Trainers</td>
<td>3365</td>
<td>3364</td>
<td>3561</td>
<td>3531</td>
<td>3683</td>
<td>3714</td>
</tr>
<tr>
<td>- in Academic Hospitals</td>
<td>1798</td>
<td>1840</td>
<td>1971</td>
<td>1995</td>
<td>2085</td>
<td>2132</td>
</tr>
<tr>
<td>- elsewhere</td>
<td>1567</td>
<td>1524</td>
<td>1590</td>
<td>1536</td>
<td>1598</td>
<td>1582</td>
</tr>
<tr>
<td>Candidate Specialists in Training</td>
<td>2374</td>
<td>2243</td>
<td>2243</td>
<td>2157</td>
<td>2181</td>
<td>2230</td>
</tr>
<tr>
<td>- in Academic Hospitals</td>
<td>1358</td>
<td>1342</td>
<td>1332</td>
<td>1291</td>
<td>1309</td>
<td>1343</td>
</tr>
<tr>
<td>- elsewhere</td>
<td>1016</td>
<td>901</td>
<td>911</td>
<td>866</td>
<td>872</td>
<td>887</td>
</tr>
<tr>
<td>Candidate Medical Specialists not in Training</td>
<td>522</td>
<td>531</td>
<td>660</td>
<td>766</td>
<td>830</td>
<td>875</td>
</tr>
<tr>
<td>Newly Registered Specialists</td>
<td>566</td>
<td>528</td>
<td>475</td>
<td>453</td>
<td>378</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 9: Annual turnover of medical specialists and entrepreneurs (x 1000)

<table>
<thead>
<tr>
<th></th>
<th>1981</th>
<th>1983</th>
<th>1985</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover (Average) of Medical Specialists</td>
<td>282</td>
<td>284</td>
<td>278</td>
</tr>
<tr>
<td>Average Income of Entrepreneurs</td>
<td>46,0</td>
<td>51,8</td>
<td>53,3</td>
</tr>
<tr>
<td>Average Income Top 10%-Entrepreneurs</td>
<td>166,9</td>
<td>177,4</td>
<td>181,8</td>
</tr>
</tbody>
</table>

82
An Economic Analysis of the (Self-)Regulation of Physicians in the Netherlands

Table 10: Number of medical trainees allowed to enter education and training

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians</td>
<td>1485</td>
<td>1100</td>
<td>1250</td>
<td>1175</td>
</tr>
<tr>
<td>- General Practitioners</td>
<td>280</td>
<td>375</td>
<td>255</td>
<td>315</td>
</tr>
<tr>
<td>- Medical Specialists</td>
<td>450</td>
<td>290</td>
<td>325</td>
<td>305</td>
</tr>
<tr>
<td>Dentists</td>
<td>150</td>
<td>205</td>
<td>235</td>
<td>220</td>
</tr>
<tr>
<td>Dentist-Specialists</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pharmacists</td>
<td>175</td>
<td>100</td>
<td>140</td>
<td>120</td>
</tr>
<tr>
<td>Obstetricians</td>
<td>75</td>
<td>75</td>
<td>30</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 11: Income of medical specialists in 1983

<table>
<thead>
<tr>
<th>Specialism</th>
<th>Income (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revalidation and Physical Therapy</td>
<td>f 86.000,-</td>
</tr>
<tr>
<td>Allergy</td>
<td>f 98.000,-</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>f 123.000,-</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>f 148.000,-</td>
</tr>
<tr>
<td>Radiodiagnostics</td>
<td>f 352.000,-</td>
</tr>
<tr>
<td>Radiology</td>
<td>f 361.000,-</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>f 366.000,-</td>
</tr>
<tr>
<td>Medical Microbiology</td>
<td>f 380.000,-</td>
</tr>
</tbody>
</table>

Table 12: Physicians: specialty, number and income (x 1000 gld)

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Income</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General Practitioners</td>
<td>6.200</td>
<td>122,3</td>
<td>1975</td>
</tr>
<tr>
<td>2. Medical Specialists, of whom:</td>
<td>9.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Anaesthesist</td>
<td>660</td>
<td>248</td>
<td>1981</td>
</tr>
<tr>
<td>3. Surgeons</td>
<td>780</td>
<td>271.9</td>
<td>1981</td>
</tr>
<tr>
<td>4. Dermatologists</td>
<td>250</td>
<td>201.9</td>
<td>1981</td>
</tr>
<tr>
<td>5. Gynaecologists</td>
<td>550</td>
<td>247.7</td>
<td>1981</td>
</tr>
<tr>
<td>6. Internal Specialists</td>
<td>1.270</td>
<td>247.6</td>
<td>1981</td>
</tr>
<tr>
<td>7. Children’s Doctors</td>
<td>600</td>
<td>155.6</td>
<td>1981</td>
</tr>
</tbody>
</table>

Table 12 (continued): Physicians: specialty, number and income (x 1000 gld)

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Income</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General Practitioners</td>
<td>6.200</td>
<td>122.3</td>
<td>1975</td>
</tr>
<tr>
<td>2. Medical Specialists, of whom:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 KNO-Specialists</td>
<td>360</td>
<td>288.9</td>
<td>1981</td>
</tr>
<tr>
<td>9 Lungspecialists</td>
<td>270</td>
<td>312.8</td>
<td>1981</td>
</tr>
<tr>
<td>10 Oculists</td>
<td>380</td>
<td>194.6</td>
<td>1981</td>
</tr>
<tr>
<td>11 Orthopedists</td>
<td>240</td>
<td>295.8</td>
<td>1981</td>
</tr>
<tr>
<td>12 Radiologists</td>
<td>560</td>
<td>353.1</td>
<td>1981</td>
</tr>
<tr>
<td>13 Urologists</td>
<td>180</td>
<td>315.1</td>
<td>1981</td>
</tr>
<tr>
<td>14 Psychiatristst</td>
<td>1.660</td>
<td>208.4</td>
<td>1981</td>
</tr>
</tbody>
</table>

Table 13: Income by form of business organization

<table>
<thead>
<tr>
<th>Form of Business Organization</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proprietorship</td>
<td>f 168.000,-</td>
</tr>
<tr>
<td>Proprietorship and Closed Corporation</td>
<td>f 307.000,-</td>
</tr>
<tr>
<td>Partnership</td>
<td>f 274.000,-</td>
</tr>
<tr>
<td>Partnership and Closed Corporation</td>
<td>f 403.000,-</td>
</tr>
<tr>
<td><strong>Average Income</strong></td>
<td><strong>Income of proprietorship</strong></td>
</tr>
<tr>
<td>Radiology</td>
<td>f 361.000,-</td>
</tr>
<tr>
<td>Chest Surgeons</td>
<td>f 320.000,-</td>
</tr>
<tr>
<td>Surgery</td>
<td>f 282.000,-</td>
</tr>
<tr>
<td>Gynaecology</td>
<td>f 271.000,-</td>
</tr>
<tr>
<td>Orthopaedy</td>
<td>f 261.000,-</td>
</tr>
<tr>
<td>Anaesthesia</td>
<td>f 256.000,-</td>
</tr>
<tr>
<td>Neurology</td>
<td>f 235.000,-</td>
</tr>
<tr>
<td>Internal Specialism</td>
<td>f 202.000,-</td>
</tr>
<tr>
<td><strong>Average Income</strong></td>
<td><strong>Income of the partnership/closed corporation</strong></td>
</tr>
<tr>
<td>Radiology</td>
<td>f 361.000,-</td>
</tr>
<tr>
<td>Radiodiagnistics</td>
<td>f 352.000,-</td>
</tr>
<tr>
<td>Pathologic Anatomy</td>
<td>f 323.000,-</td>
</tr>
<tr>
<td>Chest Surgeons</td>
<td>f 320.000,-</td>
</tr>
<tr>
<td>Surgery</td>
<td>f 282.000,-</td>
</tr>
<tr>
<td>Lungspeciality</td>
<td>f 250.000,-</td>
</tr>
</tbody>
</table>
CHAPTER 4 REVERSED SOLIDARITY IN PENSION PLANS

4.1 Introduction

An unexpected phenomenon is part of the typical Dutch pension plan. Unlike many other collective agreements such as collective labour contracts or social security arrangements, pension plans in the Netherlands increase the inequality of income distribution. In the typical pension plan, persons without a career pay for the pensions of those who have a career. This implies, generally speaking, that blue-collar workers and women pay for the pensions of white-collar workers, most of whom are men. The phenomenon raises several questions. For instance, why, in a free market economy, are contracts concluded that do not seem to be optimal from the perspective of the average employee. Furthermore, in a competitive environment, one would expect only those contracts to survive that keep the price to buyers as low as possible. What is the explanation for the survival of pension contracts that seem to be unnecessarily expensive. Finally, pension contracts are usually the outcome of collective bargaining between labour unions and employer organisations. An unequal redistribution of income seems to conflict with the alleged redistributive goal of the labour unions.

This paper is organised as follows: in the section two I will describe the general structure of pension benefits and the basic types of pension plans. In the second part of this section I analyse how non-career employees help to finance the pensions of career employees. The various aspects of this phenomenon, which I call 'reversed solidarity'\(^1\), are, for illustrative purposes, highlighted with a numerical example. In section three I will analyse the existence and stability of reversed solidarity in pension plans in the Netherlands from the point of view of 'Law and Economics'. In the fourth section I will investigate whether reversed solidarity is an international phenomenon. The paper will be concluded with a summary of the main results.

\(^1\) In the Dutch literature on pensions, the phenomenon is called ‘positive solidarity’ on account of the positive difference between actual contributions and actuarially fair contributions, See Lutjens (1989), p. 398.
4.2 The actuarial logic of reversed solidarity in pension plans

4.2.1 Pension benefits in general

Generally speaking, pension benefits in industrialised countries consist of three parts. Firstly, there is usually a state pension scheme which provides a basic pension benefit. Secondly, there are complementary pension benefits provided by the employer or by the industry concerned. Thirdly, there are personal pensions provided by the employees themselves. In the Netherlands, reversed solidarity is part of the complementary pension plans. Unless otherwise stated, the first part of this paper will deal mainly with the complementary pension plans. Furthermore, this paper only covers pension benefits for old age and not benefits related to long term disability, early retirement, surviving dependants etc.

Although complementary pension plans vary substantially across firms and industries, they are essentially of two basic kinds: defined contribution (or money purchase) plans and defined benefit plans. A defined contribution plan defines the contributions or premiums to be paid. In such a plan regular contributions from the employee and/or the employer are put into a personal account. That account is used for investment in securities. After retirement, the employee receives the accumulated amount of investment earnings and contributions in the form of a lump sum or, more often, in the form of an annuity. A defined benefit plan defines the pension benefits; the premiums and contributions are whatever is needed to provide these benefits. In a defined benefit pension plan the annual flow of pension benefits is determined by a specified formula. Although the defined benefit pension formulas show many variations and complex details, they can be divided into two broad categories: the so-called pattern or flat plans, and the conventional or unit formula plans. The formulas that determine the benefits follow below ².

The pattern or flat plans make pensions a function of the number of years of service:

\[ \text{Annual pension benefit ($)} = (\text{years of service})(\$ \text{ amount}) \]

Thus, for instance, an employee with a working life of forty years would receive $4,000 per year in pension benefits after retirement, if the benefit paid per year of service were one hundred dollars.

The conventional or unit formula plans make pension a function of years of service as well as of earnings:

\[ \text{Annual pension benefit ($)} = (\text{years of service})(\% \text{ figure})(\text{salary}) \]

There are several variants of the conventional pension plan: salary can be the (weighted) average salary of life-time employment, the average salary over the last (or highest) five or ten years of employment, simply the earnings in the final year of employment or, indeed, some other variant. The percentage figure is typically 1%-2%.

Thus, according to a conventional plan whereby retirement benefits are calculated as 2% of final earnings times number of years of service, an employee who has worked for forty years and earns $50,000 in the final year of employment would receive $40,000 per year in pension benefits after retirement.

In the Netherlands in 1985, about 78% of the employees in the private sector participated in a complementary pension plan. Of these participants, 77% had a defined benefit pension plan of the conventional type, based on final earnings and 15% had a defined benefit pension plan based on some variant of average salary. Only 7% had a flat rate pension plan and less than 1% participated in a
defined contribution pension plan\(^3\). These figures are both similar to and differ considerably from those in, for instance, the USA. Pension coverage of the work force in the USA is approximately 50\%, 75\% of whom have a defined benefit plan. But only 42\% of these 75\% have a pension plan of the conventional type and about 36\%, predominantly in the unionised sector, have a flat rate pension plan. The remaining 22\% have plans with other benefit formulas\(^4\).

### 4.2.2 The mechanism of reversed solidarity

It seems useful to clarify the mechanism of reversed solidarity by giving a numerical example of a typical pension plan. Suppose the pension plan pays a benefit equal to 1.75\% of final earnings per year of service. The plan participants enter the plan at age 25, retire at the age of 65, and live another twenty years after retirement. Consequently, retired employees will, each year, receive 70\% of their final earnings as pension benefits, which is the goal of the typical pension plan in the Netherlands. According to standard practice in the Netherlands, I assume the real interest rate to be 4\% and ignore any possible divergence between nominal and real interest rates\(^5\). Finally, I assume that the labour force consists of two equal groups of persons. In the first group (group A), employees start at a salary of $30,000 and earn an additional $10,000 every ten years. In the second group, group B, starting salary is also $30,000, but the employees have no real wage increases for the remainder of their careers. The contribution of the plan participants is specified as a uniform percentage of income, irrespective of the level of income\(^6\).

To illustrate the mechanism of reversed solidarity I shall proceed as follows:

---

\(^3\) Pensioenkamer (1989), p. 75.

\(^4\) Ellwood (1985), pp. 20-34.

\(^5\) This latter assumption does not influence outcomes significantly, since both benefits and contributions are indexed for inflation.

\(^6\) When pension premiums are interpreted as deferred payment of wages, the distinction between contributions by the employer or the employee is irrelevant in the context of reversed solidarity.
- firstly, I will determine the present value of the pension benefits at the age of retirement;
- secondly, I will determine the contribution that is required for the accrual of the pension wealth at the age of retirement;
- thirdly, I will make a comparison between the contribution as a percentage of income on the one hand and the required contributions from an actuarial point of view on the other hand.

What should the accumulated pension wealth at the age of retirement be in order to be sufficient to guarantee each person in group A $42,000 (70% of $60,000) and group B $21,000 (70% of $30,000)?

Suppose the pension benefits were given at the beginning of each year and only for three years, the accumulated pension wealth at retirement for persons in group A would be:

$$\frac{42,000}{1.04} + \frac{42,000}{1.04^2}$$

, or $42,000 \cdot (1.04)(1.04^{-1} + 1.04^{-2} + 1.04^{-3})$

Hence the present value (PV) of pension benefits for twenty years is given by the following formula:

$$PV = ($ amount)(1 + i) \sum_{n=1}^{20} (1 + i)^{-n}$$

where ‘i’ is the real interest rate and ‘n’ the number of years one is entitled to pension benefits. For each person in group A and group B, $594,000 and $297,000 respectively should be reserved for financing the pension benefits for twenty years, given the real interest rate of 4%7.

7 (1.04)(13.59) times $42,000 and $21,000 respectively.
Chapter 4

What contributions are necessary for the accrual of this pension wealth? Suppose the pension wealth were to be accumulated over three years, the following contribution for persons in group A would be required at the beginning of each year:

\[
\frac{594,000}{1.04 + 1.04^2 + 1.04^3}
\]

Thus, the required contribution from an actuarial point of view (AC) is given by the following formula:

\[
AC = (\text{\$ amount}) \frac{1}{\sum_{n=1}^{n} (1 + i)^n},
\]

where ‘i’ is the real interest rate and ‘n’ the number of years one pays the contribution.

With 4% real interest rate, each person in group B should contribute $3,005 each year for forty years, which is about 10% of his or her annual income\(^8\). Actuarially required contributions for employees in group A rise from $3,005 in the first decade to $4,701 (age 35), $6,200 (age 45) and $10,930 at the age of 55. During the last ten years of their working life, the employees in group A should contribute about 18% of their income to finance their own pension benefits. In practice, however, in the Netherlands the contribution made by the participants in complementary firm or industry pension plans is defined as a percentage of income, regardless of the level of the income. In the above example, the average contribution required for financing the pension benefits in the given circumstances rises from 10% in the first decade to 15.5% in the fourth decade\(^9\).

\(^8\) $297,000 times 0.01012.
\(^9\) Intuitively, one may have the impression that reversed solidarity may be the result of the difference of income as such, given the structure of the pension plan. Such an impression is
Although a 10% premium would be enough to pay for their pensions, employees in group B actually pay up to 15.5% of their income. At the age of 65, the value of the 'transferred' premiums is over $52,000. With that amount the annual pension benefit could be increased by almost $3,700 or 17.5%.

One could argue that, in reality, the ratio of employees in group A relative to group B is probably less than one. The consequences of the heterogeneity among employees in practice would therefore not be as large as in the numerical example above. But this assessment depends on the point of view taken. Even in the case of a small ratio of group A employees and group B employees, reversed solidarity would be significant. Given a sector where group A employees make up only 10% of the labour force while other circumstances stay the same, it is not difficult to establish that the group B employees will pay about 20% of the premiums required for the pensions of the group A employees.

misleading. Given the same pension plan and employees in group A and B earning $30,000 and $60,000 respectively, their contributions would be about 10% of their respective incomes.

10 Depending on its particular characteristics, the structure of the pension generates other forms of reversed solidarity or actuarially 'unfair' transfers:

- a complementary pension plan integrated with a state pension plan and contributions calculated as a percentage of the full wage; the contribution should take into account the state pension by way of a threshold. When the pension is 70% of final wages, the payment of pension premiums should begin as soon as the income reaches 10/7 times the state pension;
- employees with the same number of years of participation in the pension plan and the same age of retirement but different periods in which the contributions are paid. In this way their contributions generate different returns to investments;
- certain groups of employees who have a rapid increase in the earnings at the end of their career. The contributions paid would be insufficient from an actuarial point of view.
4.3 The existence and stability of reversed solidarity in pension plans

In this section I will be looking at pensions from a Law and Economics perspective. A fundamental characteristic of a competitive market economy is the freedom of contract, i.e. the freedom to determine the terms of contract and to enter into a contract with whoever one wants. The question arises as to why employees enter into pension contracts that seem to be unnecessarily expensive. Furthermore, in a competitive market economy, one would expect only those market practices to survive that keep the price to buyers as low as possible. What is the explanation for the survival of seemingly 'unfair' terms in pension contracts?

By their form, pension contracts are 'standard form' contracts. Standard form contracts are drawn up by one party only and presented to other parties on a take-it-or-leave-it basis. No negotiation about the terms of the contract is possible. Traditionally, legal scholars are rather suspicious of standard form contracts. For instance, in a well known textbook in the Netherlands one reads: 'The supplier of the standard form contract aims in the first place to promote his own interest. He has a position of economic power, sometimes even a monopoly position. The individual buyer, on the other hand, is virtually powerless.' The economic analysis of standard form contracts results in a somewhat different verdict. Standard form contracts save on search, information and transaction costs. They reduce the risk of mistakes in drawing up the contract and the risk of conflict about the content. Furthermore, they simplify the internal administration and make it easier for personnel to deal directly with customers on the basis of the contract. Standard form contracts can, therefore, be advantageous to both consumers and producers. The economic analysis of standard form contracts also specifies the circumstances and conditions that would lead to 'unbalanced' terms of the contract which favour the supplier of the contract. The major sources of this 'imbalance' in the terms

---

11 Empirical research shows that this type of contract predominates in the Netherlands, see Gras (1984), chapter 2.
of the contract are informational asymmetry, transaction costs and imperfect competition. In the next two subsections I shall investigate whether these factors might explain the existence and stability of reversed solidarity in pension plans.

### 4.3.1 Informational asymmetry and transaction cost

The acceptance of the pension contract usually coincides with entry into the labour contract. Employees, at the time, are often ill informed about the existence of the pension agreement, and are unfamiliar with the contents\(^\text{14}\). Several factors contribute to this ignorance. 

*Firstly*, when a prospective employee bargains with a firm's representative about the labour contract, the pension agreement is often not discussed. There is a simple reason for this. In the Netherlands, the law on pensions stipulates 'automatic' participation whenever the new employee belongs to a group of employees in the firm for which such an agreement exists. The employer has the legal right to deduct the necessary pension premiums from the an employee's wages to finance future pension benefits\(^\text{15}\).

*Secondly*, the nature of the pension contract precludes the acquisition of specialised knowledge about its content. The pension contract is usually executed by a pension fund. Usually these funds do not supply a copy of the (standard form) pension contract until the new employee has completed the negotiations about the labour contract and has entered the firm. Besides, since it is a standard form contract, it is not possible to change its terms. Studying the details does not create opportunities for enhancing the expected utility of the pension contract. In any case, the terms of the pension contract are difficult to interpret because of the abstract and complicated wording. The standard terms are abstract because they are intended to cover many different cases; they are complicated because of their judicial language. Particularly in the case of pension contracts, one has to make difficult and complex actuarial calculations to determine the consequences of the terms.


\(^{15}\) Art. 2, lid 2 Pensioen en Spaarfondsnew; art. 1637s B.W.
Chapter 4

Thirdly, even if new employees were to consider alternative pension contracts, they would face the formidable task of comparing, discounting and evaluating many different types of pension contracts. Most employees would find such a task too daunting. Of course, employees could acquire the necessary expertise by employing a specialist with the required knowledge, but that would be a costly business.

In the fourth place, employees have no reason to distrust the terms of the pension plan, particularly because the highly paid employees, including the management are covered by the same pension plan.

Finally, being ill informed stems from the subject matter of pension contracts and certain alleged human characteristics. For new employees, particularly the younger ones, the age of retirement is a long way off. It is often suggested that in such circumstances time preference might be distorted by myopia. As a result, future needs are systematically undervalued. Consequently, the expected utility of adjusting the pension contract would be small, whereas the cost of these adjustments would be substantial or even prohibitive. Under these circumstances, the terms of the pension contracts are likely to remain 'unbalanced'.

4.3.2 Imperfect competition

In a competitive market economy one would expect suppliers to compete on the terms on which pension benefits are offered. Any unbalanced terms would attract new suppliers in the market which would reduce or even eliminate possible inefficiencies. Unfortunately, imperfect competition results from severe limitations on the range of choices open to both the employer and the employee. Legal and institutional barriers and market failures bring these limitations about.
4.3.2.1  **Legal and institutional barriers to competition**

Whenever an employer organises a pension plan for his employees he is legally obliged to have the pension plan executed by an institution outside the firm\textsuperscript{16}. This requirement aims to render the (future) pension benefits independent of the economic welfare of the firm\textsuperscript{17}. In the Netherlands the majority of pension plans are executed by so-called industry pension funds. These funds are controlled on an equal basis by representatives of the organisations of employers and the labour unions. The terms of the pension contract are determined in the collective bargaining process between these organisations. The individual employer who is a member of the organisation of employers, is not a direct party in such negotiations. The only way to influence the terms of the pension contract is through costly initiatives and efforts within the employers’ organisation. Since the organisation is governed by majority rule, the success of such actions would be doubtful. The individual employer has, of course, the exit option to regain individual freedom of contract. But if the aim was to bargain an optimal pension contract with the employees, leaving the employers’ organisation would, in most cases, not have the desired effect. The reason for this failure is the second factor limiting the choices open to the individual employer.

When a number of employers have decided to participate in an industry pension fund, the Minister of Social Affairs and Employment has the power to declare participation in the industry pension fund compulsory\textsuperscript{18}. This obligatory participation has been effectuated for 82\% of the existing industry pension funds\textsuperscript{19}.

For the same reasons, an individual employee is not in a position to negotiate freely the terms of the pension contract. As a member of the union he is confronted with agency problems and transaction costs. If he is not a member,

\textsuperscript{16}  Pensioen- en Spaarfondsenwet, art. 2.
\textsuperscript{17}  Unlike, for instance, Germany, where firms are required to insure their pension liabilities against involvency, see Jungman (1989), p. 13.
\textsuperscript{18}  Wet betreffende verplichte deelneming in een bedrijfspensioenfonds, Wet van 17 maart 1949, Staatsblad J 121, art.3.
\textsuperscript{19}  Verzekeringskamer (1990), p. 45.
and the employer is a member of the organisation of employers who bargains over the pension contract, there is not much to negotiate\textsuperscript{20}. If neither the employee nor the employer is a member of an organisation participating in the negotiations, their freedom of contract will most likely have been rendered inoperative by the ministerial declaration to compulsory participation. Apart from these legal and institutional barriers, the functioning of the insurance market itself may restrain the options open to the employee\textsuperscript{21}.

4.3.2.2 Market failure from adverse selection and opportunistic behaviour

If the average employee were to opt for other forms of saving such as, for instance, life insurance, he would probably find the insurance policy unattractively priced as a result of adverse selection. The insurer can observe only the average probability of longevity of the insured. He will find it hard to distinguish between a high risk and a low risk policy holder and must offer the same terms to all. With perfect competition, the actuarial value of the insurance contract must be zero. This means that in a pooling equilibrium the low risk policy holder must effectively subsidise the high risk group, which results in a tendency for the high risk group to have a high demand for this kind of insurance. Furthermore, insurance companies might not even be willing to provide the desired defined benefit plans. As explained earlier, in these plans pension benefits do not vary with contributions and market returns on investments. Assume that employees can purchase life annuities at actuarially fair prices based on the career development of an average employee. The price faced by the individual employee is independent of the actual career. It could then be profitable for both the employer and the employee to deceive the insurer on the 'true' development of the career. The employer could effectively transfer some of the deferred wage payments (the pension) to the insurer; the employee would profit from the higher level of pension benefits. The insurer

\textsuperscript{20} The law declares collective agreements between the unions and the organisation of employees valid for non-union members if the employer is a member of the employers' organisation, see Wet op de collectieve arbeidsovereenkomst, art. 14.

\textsuperscript{21} Bodie (1990), p. 35, 38.
cannot monitor whether the employer and the employee conspire against him, nor can he write contingent contracts on those actions. With imperfect information and opportunistic behaviour the market might, therefore, fail to provide defined benefit pension plans. This leaves the individual employee with a choice between either an imperfectly defined benefit pension plan that is actuarially unfairly priced, or other pension plans that do not offer insurance on the level of pensions.

4.3.3 Mutual interest of unions and employers in reversed solidarity

The argument so far makes clear why any 'unbalanced' terms in pension contracts might fail to disappear in an otherwise competitive environment. Information and transaction costs are prohibitive, and collective agreements, legal requirements and market failures restrain individual freedom of choice and contract. But it is still unclear why collective pension schemes should be favourable to higher paid workers. It is usually assumed that unions alter the pension plans in ways that equalise pensions among workers. Plans for unionised workers tend to be flat benefit plans, which is the pension equivalent of standard rate policies in wages\textsuperscript{22}. It is said that those plans and policies reflect the redistributive goal of unions. If this is an accurate description of union behaviour, then what might explain the phenomenon of reversed solidarity in pension plans? One hypothesis is that the 'hidden' transfer of pension premiums is an unintended side effect of a pension system developed by actuaries and executed by functionaries of the pension funds. But this hypothesis is easily refuted by actual developments at the introduction of the pension schemes. Labour unions have been well aware of the adverse redistributive effects of defined benefit plans and ways to counterbalance these effects have been under discussion\textsuperscript{23}. In fact, the law on pensions stipulates that the management and control of industry pension funds is on the basis of parity of representation by organisations of employers and employees, which suggests that both are well

\textsuperscript{22} Bodie (1990), p. 43; Freeman (1986), p. 107.

\textsuperscript{23} Technically, the removal of reversed solidarity in pension plans is simple and straightforward.
Chapter 4

informed. Precisely this latter statutory provision might be part of the explanation of reversed solidarity in pension plans. In the given circumstances, both the organisations of employers and employees may find it in their interest to agree upon a pension plan that pays benefits on final earnings rather than on years of service.

4.3.3.1 Employers and final pay pension plans

Employers face two kinds of costs due to the specificity of the production process and the uncertainties of the employment relationship. Even in markets producing a homogenous product, firms often have different ways of combining inputs to produce outputs. As a result of the special way of organization and production, the current workforce of the firm is indispensable, at least in the short run, to maintain an efficient level of production. A longer term employment relationship develops between the worker and the firm. Important aspects of this employment relationship will be part of a contract in the form of a formal explicit document or a less well-defined implicit agreement. But in both cases at least three elements of the employment contract cannot be totally defined or agreed upon in advance:

- the actual length of the service is not determined in advance by the contract itself due to the fact that restrictions on individual mobility are viewed sceptically by the courts; consequently, the employer cannot asses whether or how long he can keep the employee or at what cost he can replace him with another;
- the exact content of the work which the employee is bound to perform is neither precisely determined nor guaranteed by the contract either;
- the intensity and quality of work is not properly measurable by the contract either.

The 'asymmetry' in information gives the employee the opportunity to work at less compared with his optimal effort. This implies that the firm could find it profitable to use compensation structures that penalize workers for poor

---

24 Pensioen- en spaarfondsenwet, art. 14.
performance, that is to dissuade them from shirking. A pension plan that pays benefits on final earnings is such a devise. If some of the employee's wages are withheld in the form of pension benefits until completion of employment there is strong incentive for employees to exert effort throughout the working life, especially if pensions are tied to final earnings. To use pensions as an incentive mechanism is also in the interest of the average employee because the value of the marginal product and thus the full wage is higher than it would be in the absence of such a compensation scheme.

But there is second explanation for pension plans that pay benefits on final earnings. These plans can be interpreted as a logical extension of the internal labour market which itself can be seen as a screening device to identify low mobility and high productivity employees for senior posts. Not all employees are equally well suited for all positions. An employee's productivity is greatest when he is matched to the right job. Furthermore, an employer will wish to match low mobility employees with jobs in which firm-specific investments are profitable. If it is too costly to acquire information about different types of employees, a firm can acquire information by experience in the internal labour market within the firm. A compensation scheme with low wages at the beginning of the career of employees and increasing wage rates with tenure makes the firm attractive for low mobility and high productivity employees. With the passage of time the firm learns more about the productivity of its employees and it rewards the most productive of these by higher earnings. Such a compensation scheme would also effectively separate employees that differ intrinsically in their job mobility propensities. Low mobility individuals would identify themselves because the compensation structure would be especially attractive for them. A pension plan that pays benefits on final earnings is a logical extension of the internal labour market as a screening device. Furthermore, it enhances the effectiveness of this device by making it costly to leave the firm.

---


26 When, for instance, an employee (of group A, see the example given above) leaves the firm after the first ten years, she will receive no more than the following pension:

\[(10 \times 1.75 \% \times $30,000) + (30 \times 1.75 \% \times $60,000) = $36,750,\] which is only 61\% of final earnings instead of 70\% if she were to stay with the firm.
There is also a simple practical reason why employers would prefer pension plans that pay benefits on final earnings. If the employer derives no productive advantage from offering pension benefits rather than wages and if the costs to the employer of providing the employees with benefits are the same as the costs of providing wages, then one would expect the employer to offer any combination of the two from which the employee can choose. However, in many countries pension funds benefit from tax deferral. Contributions are tax free, as are capital gains and interest; tax is only paid on the pension benefits after retirement. If wages are taxed, while benefits are not, this will raise the cost to the employer of offering wages as opposed to benefits. The effect of levying taxes on wages is likely to lead to a larger share of total compensation taking the form of pension benefits and a smaller proportion in wages. If other circumstances remain the same an employer would find a pension plan that pays benefits on final earnings for reasons of tax levying the superior alternative.

4.3.3.2 Unions and final pay pension plans

Despite the adverse redistributive effects of the defined benefit plans with contributions specified as an equal percentage of income, unions might also display a preference for this type of plan. As a collective democratic organisation, the union will represent the preferences of the median employee rather than of the marginal employee. The median employee will be older and closer to the date of retirement. It has been noted that rising total compensation with seniority is a widespread phenomenon in unionised firms, even for non-career employees\(^\text{27}\). Sometimes, this practice is attributed to the greater influence of senior workers over compensation policy in the union. Another analysis shows that the union might use a rising seniority-wage profile in order to extract rents from the firm\(^\text{28}\). Given the firm's downward sloping labour demand curve, efficient price discrimination by the union against the firm would require a wage profile that favours senior workers over junior workers.

\(^{27}\) Freedom and Medoff (1984), pp. 122-123.

\(^{28}\) Kuhn and Robert (1989).
In a defined benefit plan based on final earnings, most benefits are earned in the latter years of employment. Thus a preference for defined benefit plans might be explained by the median voter model of unions. The median voter, who is older and who's wages rise with seniority has much to win with a pension scheme that ties benefits on final earnings. Another view criticizes the assumption of perfect democracy in the union. The bureaucratisation of representative institutions of the union has made it possible for union leaders within certain limits to pursue their own goals. In the extreme, if the union is autocratic, the controlling group will probably be even older and thus more interested in defined benefit plans based on final earnings. Furthermore, a collective pension plan executed by an industry pension fund might be in the interest of the union leaders, since they are entitled by statutory law to occupy half the seats on the board of those pension funds29. In this way they are able to control the investments of contributions even from non-union members. Although little is known about the specific goals of the union leaders, this arrangement could further their aims in at least two ways. First, the arrangement endows the leadership with prestige and influence, which are likely to be arguments in the leadership's objective function. Second, the arrangement is likely to lead to a larger membership because workers being union members can at least exert some influence over the pension policy of the union. Not only is a larger membership in the interest of the union itself, union leadership may also be interested in having as large a union as possible rooted in the desire to maximize the dues income of the union30.

Finally, the labour union would want to eliminate competition and prevent possible downward pressures on the wage structure. Uniform pension premiums throughout the industry contribute to this end31.

29 Recall that in the Netherlands it is standard practice for the Minister of Social Affairs to use his power to declare participation in an industry pension fund compulsory once it has been established.
31 In the Netherlands, the problem of reversed solidarity is exacerbated by the typical bargaining structure. Labour organizations are split along ideological and religious lines; there are no separate organizations for blue-collar workers and white-collar workers. Moreover, the labour unions encompass various occupations, such as workers and lower
Chapter 4

To sum up, a defined benefit pension plan with contributions specified as an equal fraction of final earnings is the outcome of a bargain between the union and the organisation of employers. The employers would favour such a pension scheme because it discourages shirking, enhances productivity and the process of job matching and reduces the number of quits. Unions could also favour such a pension scheme because it is in the interest of the median voter who controls the union, because it enlarges union membership and because it enhances the prestige and influence of union leadership. Individual union-members would not oppose the specifics of the pension plan for reasons of informational asymmetry, transaction cost and the inability to capture the full benefits of their actions.  

4.4 Reversed solidarity: a Dutch peculiarity or an international phenomenon?

In this section I will investigate whether reversed solidarity as analysed above, is an international phenomenon. First of all, we shall limit our comparison to countries more or less surrounding the Netherlands: Germany, France, Belgium and the United Kingdom. With these countries the Netherlands has extensive trade relations and their societal structure is not unlike that in our country. We would therefore expect pension schemes to be of a similar structure. We shall first outline the pension schemes in the afore mentioned countries and then draw some preliminary conclusions on the structure of pension schemes and the prevalence of reversed solidarity.

---

level and middle level management, within the industry. Finally, bargaining takes place at industry level, not at enterprise level.

32 Olson (1965).
4.4.1 Outline of the pension schemes in the United Kingdom, Belgium, Germany and France

United Kingdom

The state pension scheme in The UK consist of two parts. The first part is a 'flat rate' basic pension. The basic old age pension is £3,900 for a married couple and £2,450 for a single person (April 1990). The pension is payable to all with a sufficient contribution record. The second part is an additional state pension based on earnings and length of employment. This state earnings related pension scheme (SERPS) is compulsory for all employed persons with earnings between a lower earnings limit and an upper earnings limit. The lower earnings limit is equal to the level of the basic state pension, the upper limit seven times that. The SERPS pension is currently, at most, 25% of the employees best twenty years re-valued earnings (within the income limits), the pension being reduced to 20% of average lifetime earnings by the end of the century. Pension premiums are a percentage of income, ranging from 5% to 9% for both the employer and the employee, depending on the level of income of the employee. These premiums are used for direct financing of current pension benefits (‘pay-as-you-go’). Although the SERPS scheme is compulsory, it is possible to contract out of SERPS. Until 1988 this option was open to a decision made by the employer only (though following consultation with the employees). The SERPS pension had to be substituted by a defined (minimum) benefit pension, with roughly the same benefits the employee would have got if he had stayed in SERPS. Since April 1988, the employee has a choice of whether to stay in SERPS, or to join the employer's pension plan or to take out an appropriate

33 Cunliffe (1990), pp. 78-80.
35 Financing on a pay-as-you-go basis implies that employees in year t pay for the pensions of the retired in year t. In such a system of pension financing, any form of equivalence of between the payment of pension premiums and the pension is lacking. Consequently, transfers are between generations of employees and the retired, and the intragenerational phenomenon of reversed solidarity does not occur.
personal pension. Since then it has also become possible to opt out of the SERPS pension for a defined contribution pension plan, the minimum contribution being 5.8% of earnings between the lower and upper earnings limit. Complementary to the state pension are the so called occupational pension plans. These are contractual arrangements between the employer and the employee or arrangements in which the employer effects an insurance policy or pays money to trustees who pay the benefits. Most of these plans are defined benefit plans and most plans were used to contract out of SERPS. Defined contribution plans will probably become more important since younger employees profit from the non-age related minimum contribution required to contract out of SERPS.

About 55% of all employees participate in an occupational pension plan, although there is no legal requirement on an employer to establish a pension plan. Since 1988 it is illegal to make membership of pension plans compulsory. Before that time the decision to enter the pension plan was up to the employer and occasionally the result of collective bargaining.\(^\text{36}\)

**Belgium**

As in other countries, the structure of the pension scheme in Belgium consists of at least two parts: the state pension plan and some complementary plan organised by the employer.

The magnitude of the state pension benefit depends on contributions and the years of service. Contributions are compulsory for the employed. The contributions (pension premiums) are 8.65% of earned income for the employee and 11.06% for the employer.\(^\text{37}\) The maximum annual income on which premiums are levied is BF 1,211,00. These premiums are not used for investment purposes but for direct financing of current pension benefits (pay-as-you-go system).

In principle, the retirement pension amounts to 75% (married couple) or 60% (single person) of average earnings, limited to certain salary ceilings. These maximum pensions can be obtained after 45 years of employment; in February 1990 they amounted to BF 568,061 (married couple) and BF 454,449 (single


person). There is also a minimum pension benefit to be obtained after 45 years of employment: BF 352,015 (married couple) and BF 281,649 (single person). The complementary pension plan organised by the employer is executed either by a pension fund or by an insurance company. There is no legal obligation to join the pension plan, but the plan is considered to be part of the overall working agreement between the employer and the employee. When a pension plan exists, new employees will, in principle, be obliged to join the plan. Approximately 30% of employees participates in a complementary pension plan. This comparatively small percentage is due to the rather large state pension benefits. Only the higher paid employees need additional pension benefits to maintain their standard of living after retirement. Premiums are invested in securities and these investments and their returns are used for the payment of the pension benefits. Defined contribution systems (for instance, a contribution of 2% of the wage for the employee and 4% for the employer) used to predominate, but pension contracts specifying pension benefits are becoming increasingly important. They usually aim at maintaining the standard of living after retirement, by specifying pension benefits as a percentage of final earnings (most of the time including the state pension plan).

Germany

The state pension scheme provides retirement pensions for wage earners and employees on a salary. Both the employee and the employer pay a 9.35% pension premium up to an annual income of DM 78,000. Pension premiums are used to finance current pension benefits (pay-as-you-go system). Pension benefits to be received depend on income and on years of service. The income concerned is an adjusted average income. Maximum benefits are received after 40 years of employment; each year of service generates up to 1.5% of the corrected basic income. The individual correction factor depends on personal income in relation to average income of all members of the pension

Chapter 4

plan, per year of participation. With the maximum years of service the pension received amounts to about 64% of average earnings (after tax).

The complementary pension scheme in Germany is rather complicated. Generally speaking, there are four types of complementary pension schemes. These types can be combined in one and the same company pension plan. About 52% of all employees participating in a pension plan (60% of all employees) will benefit from the so called 'Direktzusagen' by the employer: written promises to pay pension benefits after retirement. The employer finances the benefits from the 'book reserve': funds set aside within the firm. In this case usually only the employer pays the necessary contribution to meet his pension obligations. The employer is obligated to reinsure the pension promises against insolvency. The second type of complementary pension schemes takes the form of an individual or collective insurance contract, taken out by the employer on behalf of his employees. In this case, additional contributions by the employees are possible. Twelve percent of all participants in a pension plan will benefit from such a direct insurance contract ('Direktversicherung'). Another 25% will obtain pension benefits from the so called relief funds ('Unterstutzungskassen'). These funds, which originally developed from charity, can be described as legal entities that pay pension benefits but grant no title to such benefits. Participants cannot contribute individually to these funds. Finally, nine percent of all participants in a pension plan will receive their pension benefits from a pension fund. The future benefits are funded by contributions made by the employer, and sometimes also by employee contributions.

There is no legal obligation to join a pension scheme, but such an obligation can be created on the basis of a collective agreement with workers' councils (they are seldom part of wage agreements with trade unions). Because of the different forms of pension schemes, it is difficult to make general statements about the character of the promised benefit: whether it is a defined benefit or a pension fixed by the amount of the premium paid. But in most cases it is the employer who pays the pension premiums. For much the same reason it is difficult to make general statements on the relationship between the level of pensions and the earnings of the employee. Generally speaking, one can only say that most

---

41 Pauly and Heubeck (1989); Petersen (1989).
42 Idem, pp. 59-72.
pension plans offer a fixed amount dependent largely on the duration of employment and that, in practice (1990), the average pension benefit is about DM 4,200 a year.

France

The state pension scheme in France provides a pension, the amount of which depends on salary and number of years of service. Maximum pension is received after 37.5 years (150 quarters) of service (and contributions). The pension is 50% of the best ten years of average earnings up to a salary ceiling of FF 137,760, which is the social security ceiling\(^{43}\). The maximum state pension is thus FF 68,880. There is also a minimum state pension (about FF 30,000). The state pension is increased by 10% when the employee has brought up three or more children. In practice (1990), a retired person receives about 45% of his salary, restricted to the ceiling. Pension premiums are a percentage of income: 6.6% for the employee and 8.2% for the employer. These contributions are used to finance the benefits of current pensioners ("pay-as-you-go"). Since 1972, the complementary pension plans have become compulsory for all employees in industry, commerce and the private sector in France. There are three separate types of pension plans: for management and employees on a salary (Cadres), for wage earners (Non-Cadres) and for senior management (Cadres Superieur). The pension plan for senior management is not compulsory and is of a supplementary nature. In the other two plans, pension premiums vary according to income. There are three income brackets for cadres-employees: up to the social security ceiling, between that income level and four times the social security ceiling and above that. The (minimum) premium percentages are about 2%, 4% and 7% respectively for the employee, and 3%, 9% and 11% respectively for the employer\(^{44}\). For the non-cadres employees, the pension premium is about 2% (and 3% for the employer) up to a salary of three times the social security ceiling. Additional voluntary contribution of about 50% of the minimum required premium percentage is possible. These premiums are used to finance benefits for current pensioners ("pay-as-you-go"). The acquired

\(^{43}\) Stackler (1991), pp. 50-53.

\(^{44}\) Petersen (1989), pp. 65-90.
entitlements on pension benefits are expressed in units called 'points'. The total amount of pension benefits is the sum total of points multiplied by the value of the points. The value of the point is assessed periodically. In practice, after 40 years of employment, non-cadres employees can expect a retirement pension of about 65 to 75% of final earnings (including the state pension). For cadres employees, this percentage ranges from 40 to 60% of final earnings.

4.4.2 Preliminary conclusions on the international prevalence of reversed solidarity in pension plans

What is striking at first sight, unlike what we expected, is the complexity and heterogeneity of the pension schemes in the countries observed. We shall make some further comments on this heterogeneity, but first we shall draw some conclusions on the prevalence of reversed solidarity in the countries surrounding the Netherlands.

First, the scale of the state pension provision differs considerably. The replacement ratio\textsuperscript{45} of the state pension provision (average state pension as a proportion of average earnings) is low, for instance, in France and high in Belgium. A high replacement ratio effectively limits the scope for complementary pension schemes. Since state pension schemes are invariably of the defined benefit type using a pay-as-you-go system for financing, reversed solidarity is of less importance in countries with a high state pension replacement ratio. We would, therefore, expect reversed solidarity to be of lesser importance in Belgium and perhaps Germany.

Next, the accumulation of funds by or on behalf of employees to finance their pensions is, in some countries, more important than in others (see table 4.1).

\textsuperscript{45} Davis (1991).
Table 4.1  Investment in securities and property by pension funds as a percentage of gross domestic product (1989)

<table>
<thead>
<tr>
<th></th>
<th>Netherlands</th>
<th>Germany</th>
<th>Belgium</th>
<th>France</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>76%</td>
<td>4%</td>
<td>5%</td>
<td>3%</td>
<td>47%</td>
</tr>
</tbody>
</table>


In France, complementary pension schemes are an unusual combination of defined contributions for entitlements on pension benefits and pay-as-you-go funds for financing the pension benefits. Both features rule out reversed solidarity. In Germany, the majority of employees do not contribute (directly) to complementary pension plans. Pensions have their basis in the benefit pledge and pension benefits are financed from the book reserve. In Belgium, the pension plans used to be of the defined contribution type, which, in general, excludes reversed solidarity. However, since defined benefit types are becoming increasingly popular, further research into the scope and the particulars of these plans should make it possible to draw more general conclusions. In the UK, the capitalization principle and defined benefits based on final pay underlie the complementary pension schemes. In so far as the pension scheme is the result of contracting out of SERPS, differentiation of pension premiums according to classes of income might mitigate the degree of reversed solidarity. Once again, further research into complementary pension schemes, not being used for opting out of SERPS, should make it possible to draw more general conclusions.

46 In defined contribution plans there is a strong equivalence between the contribution paid and the pension benefits, reversed solidarity is a phenomenon connected with defined benefit plans, and especially the final earning version of these plans.
Chapter 4

From a 'law and economics' point of view the heterogeneity and diversity of the pension schemes raises several important positive and normative questions:
- Questions of a positive nature are: how can we explain the fundamentally different pension structures, even between countries that have extensive trade relations and comparable societal structures? Why is it, that is some countries pensions are financed on a pay-as-you-go basis, while other countries cover their pensions by yields from capital investment. Why do find in some countries pension schemes of the defined contribution type and in other countries defined benefit pension schemes? What are the incentive effects of the different type of pension schemes and are these differences important from an 'international competitiveness' point of view? Why is that in some countries unions control pension funds, while in other countries employers determine the content of pension plans? Why are in some countries pension plans part of the wage agreements with trade unions, while in other countries they are not?
- Questions of a normative nature are: is it possible to determine an optimal structure of pension schemes and what are the determining variables. Are there any market failures and how should be dealt with them: are employees imperfectly informed on future preferences and financial possibilities, and if so, what is the optimal form of regulation. Are there perhaps market failures in the financial market itself, which make it necessary to interfere with individual decision making or are external effects on the system of social security the main reason for governmental regulation? What is the optimal level of decision-making on pension plans: is it the individual level, the sector level or the national or international level?
So far, these question have hardly been addressed in the law and economics literature: new research programmes still have to be developed.

4.5 Summary and conclusions

There is a unexpected phenomenon in the majority of complementary pension plans in the Netherlands. Unlike other arrangements that are the result of collective bargaining and decision making, these pension plans increase the inequality in the distribution of (lifetime) income. In those plans, persons without a career contribute to the payment of pension provisions of those with
a career. Generally speaking, this implies that blue-collar workers and women pay for the pensions of white-collar workers, who are mostly men. For some of the contracting parties, the terms of the pension contract seem to be disadvantageous and sub-optimal. The question arises as to why these contracts are being concluded and how they can survive in a competitive environment. Moreover, since the pension plans are the result of collective bargaining between the organisation of employers and the labour unions, the question arises as to how reversed solidarity fits the alleged redistributive goal of the labour unions. The analysis leads to the following conclusions. Firstly, that information and transaction cost, collective agreements, legal barriers and market failures on substitute arrangements prevent the conclusion of optimal contracts. Secondly, employers would find the pension plan attractive because it discourages shirking, enhances productivity and the process of job matching and reduces labour turnover. Thirdly, a pension plan based on final salary is in the interest of the median voter in the labour union, who is older and earns a higher seniority wage than junior workers. Furthermore, in the given circumstances the pension plan enlarges union membership and the dues income of the union and endows the union leadership with more prestige and influence. Individual union members would not oppose such a pension plan for reasons of informational asymmetry, transaction cost and the inability to capture the full benefits of their actions.

A comparison of the pension schemes of Belgium, Germany, France and the UK shows that there are large differences in the structure and content of retirement provisions. Reversed solidarity may also be part of the pension plans in Belgium and the UK. The 'law and economics' literature on the explanation of pension schemes, their effects and the determination of optimal pensions needs further development.

Reversed solidarity results from pensions based on the capitalization principle, of the defined benefit type, with pension benefits depending on final earnings and contributions specified as an equal percentage of income.

Acknowledgement
The author is grateful for the valuable comments made by C.K.F. Nieuwenburg, J.J. Siegers and two anonymous referees, none of whom may be held responsible for the views expressed in this paper.
CHAPTER 5  NONCOMPETITION CLAUSES: UNREASONABLE OR EFFICIENT?

5.1 Introduction

Noncompetition clauses, which are also called ‘covenants not to compete’ or ‘restrictive covenants’, are encountered in many legal contexts. These clauses are often included in agreements to sell a business and in employment contracts. This paper will concentrate on the latter type of contract. A noncompetition clause protects the employer from competition by the employee after his employment has ended. For instance, the agreement may oblige the employee not to work for a competitor in the same region, or for an employer in the same sector, product market or line of work for a certain period of time after the employment has ended.

In many countries, legislatures and courts have felt the need to protect employees against these clauses. In some countries or states, such as Mexico, Uruguay, California, Alabama and Oklahoma, contractual restraints on engaging in a lawful profession are simply forbidden\(^1\). In other countries or states, such as Belgium, Germany and Spain, Texas and Michigan, legislators have restricted the use of noncompetition clauses by statute. To be valid in these countries and states, noncompetition clauses have to comply with certain legal standards laid down in law. These legal standards may refer to limitations with respect to the duration of the clause in the post-employment period (one to five years), the nature of the job (only allowable for higher paid employees), the region where the employer or employee is active, and so forth. The law may even declare noncompetition clauses valid only if post-employment payments have been agreed in exchange for promises not to compete after the employment has ended. In Belgium and Germany, for example, the employee will have to be provided with at least half his former gross wage for the duration of the noncompetition period. In some countries (Belgium, France) collective bargaining agreements

\(^1\) Blanpain, Richey and Malsberger (1991).
may impose limitations on the scope of individual noncompetition agreements. In countries such as the Netherlands and The UK as well as in the majority of the American states, however, there are no material statutory restrictions on the use of noncompetition agreements. At most, as in the Netherlands, the law may specify some formal conditions for noncompetition agreements: they have to be written and agreed by employees who are of age. In countries without material statutory restrictions on noncompetition agreements, it is left to the courts in cases of dispute to determine whether the clauses in the agreements are valid or not. In these cases, judges have generally acknowledged that employers and former employees have conflicting interests. Employees will want to make full use of their knowledge and experience but employers have a legitimate interest in protecting their trade secrets and confidential information. Whether in the United States or in the Netherlands, courts will declare noncompetition clauses valid if they are reasonable, and the clauses are reasonable if the conflicting interests are well-balanced. The courts perform a reasonableness analysis to judge whether the interests are well-balanced. In the US, a noncompetition agreement will be declared valid “if it is no greater than required for the protection of the employer, does not impose undue hardship on the employee and is not injurious to the public”. Similar wordings can be found in other jurisdictions where courts have to decide on the reasonableness of noncompetition clauses.

As this short survey suggests, a noncompetition agreement, although the result of free bargaining and contracting, is generally looked upon unfavourably by both legislators and judges. They both see reason to deviate in their decisions from clauses agreed by employers and employees. Both legislators and judges feel the need to protect employees against unbalanced clauses and they take their decisions accordingly. But why would employees enter into contracts if

---

3 In the Netherlands, an amendment of the law on noncompetition clauses is in preparation. The purpose of the amendment is to lay down in law a number of material restrictions on noncompetition clauses.
these contracts would not be in their interest? In this paper I will evaluate the different theories advanced on noncompetition agreements. Firstly, I will analyse the traditional legal theory to intervene in the freedom of contract on behalf of the employees. Are noncompetition clauses unreasonable almost by definition because employees have an inferior bargaining position? Secondly, I will review within the framework of economic theory a number of other arguments put forward by legislators and judges in their evaluations of noncompetition agreements. Does economic theory confirm the prediction that noncompetition clauses serve to protect trade secrets and confidential information? And is it true that we may expect an employer to use noncompetition clauses as a device to obtain a refund of his expenditure on ‘on-the-job’ training, as is often argued by legislators and judges? More precisely, I will analyse the circumstances under which noncompetition clauses may be predicted from the following theories: the economic theory of self-selection and screening, human capital theory, the economic theory of intellectual property and the economic theory of incomplete contracting. I will conclude that, contrary to what is often thought, it is not to be expected that noncompetition clauses generally arise from the need of employers to have their expenditure on training refunded. A noncompetition agreement can be better explained as a device to restrict opportunistic behaviour and hold-up problems, an argument not found in the legal discussions on noncompetition clauses. Furthermore, noncompetition agreements may serve to protect trade secrets and confidential information, but it is questionable whether this implies that such agreements are also in the interests of society at large.

5.2 Traditional legal arguments to restrict noncompetition clauses

Why do employees enter into contracts that impose undue hardship on them? The unequal bargaining position of the employees is at the heart of the traditional legal theory on noncompetition agreements. Because employees have an unequal bargaining position, they need to be protected, and contracts may need to be modified in order to be reasonable. The unequal bargaining position results from inequalities in the distribution of wealth. An owner of capital may diversify his risks and invest his resources in different sectors. The income of an
Chapter 5

owner/employer is thus independent of the returns on the investments in a specific sector. If the returns are insufficient, the owner can simply withdraw his capital and invest his resources in another sector. A worker, on the other hand, does not own capital, and borrowing is not an alternative because of possible bankruptcy, among other reasons. The worker is thus dependent on the owner/employer for his income. Since the owner/employer has several options and the worker has none, freedom of contract is illusory. Furthermore, it is often argued that workers have insufficient information on the noncompetition agreement and its consequences. And even if they were to succeed in obtaining this information, it would not change their bargaining position because the employment contract and the ancillary noncompetition agreement are for the most part standard form contracts that are non-negotiable. A non-negotiable standard form contract offered by an employer can be expected to be in the interest of the supplier alone, and for that reason it is almost by definition an unbalanced contract⁵.

The agreement not to compete appears to be detrimental to the employee. If he wants to apply for a comparable position with a competing employer, he has a choice of whether or not to consult his current employer. By consulting his current employer at an early stage of the application process, an employee could be putting his future career at stake. The employee could also wait until he is sure of obtaining alternative employment and then consult the employer. If the employer insists that the agreement is kept, the employee has no option but to try to have the clauses declared invalid by the courts. Courts only declare agreements unlawful if they find excessive restrictions, but even then they may try to save the agreement by modifying an excessive area restriction or an excessive time restriction, for example. The employee could also breach the noncompetition agreement and enter into an employment contract with a subsequent employer. The former employer may then have the agreement enforced and a court will issue an injunction that requires the employee to refrain from com-

---

⁵ This view on standard form contracts is expressed by the Dutch legislature, see the introduction of the new Civil Code on standard form contracts (books, 3, 5 and 6 of the Nieuw Burgerlijk Wetboek: Algemene voorwaarden), Kamerstukken I 1986/1987, nr. 16 983, p. 1.
peting against his former employer. The employee is then out of a job. Furthermore, if there is a liquidated damages clause included in the noncompetition agreement, the employee will have to pay, possibly considerable, damages as agreed at the time of concluding the employment contract. Finally, he may also lose post-employment benefits linked with the agreement not to compete, such as retirement benefits. All these consequences of the noncompetition agreement appear to be a great restraint on, and damaging to, the employee.

The foregoing analysis leaves a number of questions open, however. These questions concern the role of search and information in relation to competition, the implications of the competition among employers for employees and the probability of unfair terms in standard form contracts. First of all, assume there are two categories of employers offering an employment contract. The two categories differ only in that the one offers employment contracts without noncompetition clauses while the other includes these clauses in their contracts. All other relevant aspects of the employment contracts are identical. A utility-maximizing employee will, of course, enter into a contract without the clause ‘not to compete’. Employers in the other category will find it difficult to fill their vacancies and they will have to increase wages or non-wage benefits to compensate for the noncompetition clause in the employment contract. It might now be argued that employees are not informed about the terms of the employment contract: who would read, let alone understand, the fine print in the employment contract. However, in those employment relationships where covenants not to compete appear to be significant, it is unlikely that potential employees would be unaware or uninformed of their consequences. Covenants not to compete are not uncommon in contracts for designers, research and development personnel, salesmen, computer system personnel, managers and the like. An analysis of all court cases involving noncompetition clauses in the Netherlands up to 1995 also shows that the great majority of the cases concerned higher level managers, middle-level executives and salespersons (see table 5.1 below). We would expect these employees to be sufficiently qualified and experienced to understand the meaning and consequences of noncompetition clauses.
Table 5.1: Positions and noncompetition court cases in The Netherlands

<table>
<thead>
<tr>
<th>Position</th>
<th>Percentage*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Top-level management</td>
<td>6.8%</td>
</tr>
<tr>
<td>2. Middle-level management</td>
<td>36.2%</td>
</tr>
<tr>
<td>3. Highly educated executives</td>
<td>15.5%</td>
</tr>
<tr>
<td>4. Professional, consultancy services etc.</td>
<td>11.4%</td>
</tr>
<tr>
<td>5. Sales and acquisitions</td>
<td>32.7%</td>
</tr>
<tr>
<td>6. Special client-related services</td>
<td>5.7%</td>
</tr>
<tr>
<td>7. Standard administrative and manufacturing occupations</td>
<td>13.2%</td>
</tr>
</tbody>
</table>


*) Some cases were classified into more than one category.

Furthermore, competing employers offering employment contracts without noncompetition clauses will emphasize the comparatively favourable terms of the contract. Moreover, not every employee has to be informed of the noncompetition clauses. A small group of informed employees who will search for the best employment contract, will motivate an employer to adapt the terms of the employment contract to that of the competing employers. The tendency towards equalization of contract terms will take place when the costs of being unable to contract the informed employees rises above the benefits of using the noncompetition agreements. The competitive process assumes that some employees will read and understand contracts. Employees will search for better contracts when the expected costs of the search are outweighed by the expected benefits. The expected benefits of the search will depend on the expected value of the employment contract and the differences in the expected value of employment contracts among different employers. An employment contract will

---

7 Gazal (1999).
have a higher value than, for instance, a contract for a ‘one time bus ride’ and employees will therefore read employment contracts more readily than passengers read transport contracts. On the other hand, the noncompetition agreement may only be a small part of the total value of the employment contract: it is the present value of a future situation with an uncertain expected value. Furthermore, if the probability of entering into an employment contract with unfavourable terms is small, so that the expected benefits of the search are small, then some employers may escape the shake-out of the competitive process. It is therefore possible that some employers with unfavourable contract terms will be sheltered from competitors and survive the process of competition. Unfavourable contract terms can also be expected from ‘fly-by-night operators’, but those businesses will not include noncompetition clauses in their employment contracts for obvious reasons.

Secondly, legal scholars often interpret the terms of noncompetition clauses as unfair and explain these terms from the monopoly position of the employer. The employee has no choice but to accept the unfavourable terms of the contract. Not so long ago, a dominant approach among economists was to explain incomprehensible practices as monopolistic behaviour. As recently as 1972, Ronald Coase wrote “…….if an economist finds something - a business practice of one sort or other - that he does not understand, he looks for a monopoly explanation”\(^8\). But even in monopolized markets, we do not find noncompetition clauses attached to all employment contracts. A monopolist employer does not automatically include unfavourable terms in employment contracts. Favourable terms in the employment contract could lower the reservation wage and profits could increase if the costs of these better terms are lower than the benefits in the form of lower wages. It is, of course, still true that in the absence of unions the terms in employment contracts in monopolized labour markets will be worse than comparable contract terms in competitive labour markets. Either wages will be lower or the remaining terms of the employment contract will be worse. But there is no empirical evidence that economies abound with monopoly labour markets in all sectors.

---

\(^8\) Coase (1972).
Table 5.2: Businesses with noncompetition clauses as a percentage of all businesses

<table>
<thead>
<tr>
<th>Sector</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Agriculture and Fishery</td>
<td>27%</td>
</tr>
<tr>
<td>2. Industry and Mining</td>
<td>35%</td>
</tr>
<tr>
<td>3. Construction</td>
<td>15%</td>
</tr>
<tr>
<td>4. Wholesale Trade and Retail Trade</td>
<td>33%</td>
</tr>
<tr>
<td>5. Catering</td>
<td>3%</td>
</tr>
<tr>
<td>6. Transport and Communications</td>
<td>41%</td>
</tr>
<tr>
<td>7. Finance, Insurance and Business Services</td>
<td>40%</td>
</tr>
<tr>
<td>8. Social and Community Services</td>
<td>20%</td>
</tr>
<tr>
<td>9. Services not classified elsewhere</td>
<td>64%</td>
</tr>
<tr>
<td>Total</td>
<td>34%</td>
</tr>
</tbody>
</table>


Table 5.2 illustrates the fact that noncompetition clauses are not included in employment contracts for all firms in a sector. For instance, we find noncompetition clauses in 35% of the businesses in industry and in 3% of the businesses in the catering sector. From these figures, it would appear that employees could effectively search for employment contracts that do not include the challenged covenants. It has only been since the early seventies that economists, instead of relying on monopoly explanations, have tried to find efficiency explanations for unusual terms in contracts and forms of organizations.

Thirdly, are standard form contracts almost by definition disadvantageous to those who enter into those contracts? Several arguments have been presented in legal and economics literature that contradict this assumption commonly made by legal scholars\(^9\). Employers will compete among each other to have their va-

---

\(^9\) Katz (1998); Trebilcock and Dewees (1981).
cancies filled and standard form contracts will develop in much the same way as one price tends to develop in the market. Furthermore, standard form contracts lower the cost of search for employees and they save on time-consuming and costly negotiations. Standard form contracts also lower transaction costs in that they simplify the internal administration, reduce the probability of making mistakes and make it easier to decentralize decision making in the firm. They make markets more transparent, enhance competition and improve the efficiency of markets. Of course, as argued before, the marginal employee must have an incentive to read contracts, search costs must not be prohibitive and labour markets should not be monopolized. Finally, it might be argued that non-competition agreements generally result from incompleteness of information and transactions costs. Employers systematically underestimate the (implicit) costs of these clauses and routinely incorporate them in their employment agreements, whereas employees systematically underestimate the readiness of employers to enforce these clauses or overestimate the willingness of the employer to waive the agreement. However, there is no empirical evidence that these assumptions are generally satisfied. We must therefore conclude that non-competition clauses generally are not unfair in the sense that they are not effectively disciplined by market forces, but that under special circumstances they may be unbalanced. Nevertheless, one in every five employees in the Netherlands enters into a noncompetition agreement\textsuperscript{10}. The question then still remains as to why employees enter into such seemingly disadvantageous agreements.

5.3 Economic explanations of noncompetition clauses

Legislatures and courts generally recognize the interests of the employer in including covenants not to compete in their employment contracts. In this third part I will evaluate four theories that may explain noncompetition clauses in employment contracts. If noncompetition clauses are efficient, they enhance the value of the employment contract. This would mean that such clauses are in the interest of both the employer and the employee. It is often argued in the legal

\textsuperscript{10} Van der Werf, Ziegelaar and Schoenmakers (1997).
Chapter 5

literature that employers include noncompetition clauses to protect their investments in on-the-job training and to protect their trade secrets and confidential information. After scrutinizing the arguments I will develop two additional explanations for noncompetition clauses from the economic theory of employment relationships. Noncompetition clauses may be explained from screening and selection difficulties and from the risk of opportunistic behaviour in incomplete long term employment contracts.

5.3.1 Noncompetition clauses to refund training expenditure

It is often argued that noncompetition clauses are instrumental to the protection of investments in ‘on-the-job’ training by employers\textsuperscript{11}. If noncompetition clauses were to be modified or declared invalid this protection would no longer be available and investments in training of the employees would be reduced. To determine whether noncompetition clauses plausibly protect these investments, a short digression on general and specific ‘on-the-job’ training is necessary.

The distinction between general and specific training was developed by Becker (1975) and their connection with noncompetition clauses was first introduced by Rubin and Shedd (1981). The distinction is relevant for the predicted distribution of costs and benefits of both kinds of ‘on-the-job’ training. Specific training is training that increases the productivity of the employee only with his current employer, while general training increases his productivity with all future employers. Training for a firm-specific computer program is specific training, while training in the use of Microsoft Word is general training. While she is being trained, the training of an employee will at first diminish her productivity, but her productivity will increase after she has completed the training. Training is costly and a firm will only invest in training if it creates a surplus of benefits over costs. These benefits will materialize if the employee accepts a wage lower than her marginal product after training. However, if the training is of the general kind, other employers will be willing to pay the trained employee accord-

\textsuperscript{11} Rubin and Shedd (1981); Reitsma (1999); Shadowen and Voytek (1984).
ing to her actual marginal product and that will induce the employee to leave the firm after training. On the other hand, if the training was of the specific kind, then the employee could be persuaded to stay with the firm for a wage slightly higher than her wage with other employers. Training costs could of course also completely be borne by the employees through a temporary reduction in wages. But employees would fear being fired after getting paid a wage lower than their productivity and thus lower than the market wage. If both the employer and the employee were to pay for the costs of firm-specific training, it would be in the interest of both to continue the employment relationship. Employers will recoup their investments because employees will stay with the firm while employees need not fear being fired and will be paid a wage higher than the market wage. The distinction between general and specific training thus leads to the conclusion that, in the absence of transaction costs, employees will pay for their general training and the employer or both the employer and the employee will pay for the costs of specific training.

What is the role of noncompetition clauses with respect to ‘on-the-job’ investments? It is argued in the legal literature that noncompetition clauses protect the investment of the employer by inducing the employee to stay with the firm. First of all, one may notice that if a noncompetition clause is to protect the employers’ training investments, the offer to invest in the employee will have to be part of the employment contract. Otherwise, after having agreed with the noncompetition clause, an employee would be vulnerable to opportunistic behaviour by the employer, in this case a lower wage without the training. The employee will anticipate this risk and demand guarantees in exchange for his promise not to compete. If the employer offered general training, a noncompetition clause may protect the investments. However, to pursue this goal, a noncompetition clause would be a disproportionate and costly instrument. For an employee, the costs of such a clause would be uncertain and, since employees are often assumed to be risk-averse, she will demand full compensation in re-

12 If mobility costs, such as the psychological costs of leaving a familiar environment, or monetary relocation costs, are sufficiently high, the employee may not leave the firm even if the market wage is higher than her current wage.
turn for the inclusion of the clause\textsuperscript{13}. The compensation would include a price for the acceptance of the uncertainty of the nature of future jobs, for the restriction of the choice of future jobs and for the expected decline in future wages. For an employer it would be less costly to simply add a clause into the employment contract requiring the employee to pay back the cost of the training if the employee leaves the firm within, say, two years of completing it. If the employer offered \textit{specific} training, the employee’s productivity increases only with the firm she is currently employed by. In this case the employee’s future wages would decline rather than increase in the event of resigning and seeking work elsewhere. There is some indication that employers do not see noncompetition clauses as a means of protecting training investments. In the Netherlands, a quarter of all firms that include noncompetition clauses in their employment contracts also include a clause that requires employees to pay back training costs if they leave the firm soon after completing the training\textsuperscript{14}. It would appear that the refund of investment expenditure is not the main reason why employers include noncompetition clauses in their employment contracts.

5.3.2 Noncompetition clauses to protect trade secrets and confidential information

In the legal literature it is commonly assumed that noncompetition clauses serve to protect trade secrets and confidential information. Trade secrets and confidential information are broad and sometimes overlapping concepts. In both cases it concerns information that gives the possessor a competitive advantage. Trade secrets may consist, for example, of marketing strategies, chemical formulas, working methods, a product composition, specialized machinery and so forth. Confidential information may concern diverse objects such as customer lists, inside information of the firm, quality ratings of raw material or

\textsuperscript{13} In a number of countries (for example Germany, Belgium, Spain, Portugal, most American states) it is not uncommon for an employee to receive her salary or a part thereof for the duration of the noncompetition period.

\textsuperscript{14} Van der Werf, Ziegelaar and Schoenmakers (1997).
other inputs in the production process etc\textsuperscript{15}. The economic problem with the production and distribution of information is well known. It is often difficult to prevent others, including former employees, from using the information, and at the same time the production of information is costly. If producers do not succeed in appropriating the full value of the information they produced, economic theory predicts that it will be under-produced. One solution to this problem is the creation and protection of property rights in information, such as a patent. However, a producer will often fail to satisfy the criteria of patent law, such as novelty, practical utility and non-obviousness, or he simply wants to save the costs of establishing such a right. Another solution would be to include a non-disclosure covenant in the employment contract. In a nondisclosure covenant an employee promises not to disclose secret information to competitors for a certain period after termination of employment with his current employer. These covenants are common practice. In the Netherlands, 74\% of employers who make use of covenants not to compete also include such secrecy clauses in their employment contracts\textsuperscript{16}. One problem with these clauses, however, is that it can be very difficult to ascertain if the agreement not to disclose has been breached, and it would be even more difficult to prove it\textsuperscript{17}. Yet another solution for the employer is to resort to tort law. Liability is sometimes laid down in separate statutes such as a law of trade secrets misappropriation. However, use of an injunctive relief in trade secret misappropriation, requires proof of actual or inevitable use of confidential information. Furthermore, trade secrets cases also require proof of damages resulting from disclosure of secret information by former employees to third parties. Given these difficulties and depending on the amount of compensatory payments employers have to pay to include non-competition clauses, a covenant not to compete may be an efficient solution for the protection of proprietary information.

\textsuperscript{15} Generally speaking, trade secrets more often concerns information about the production process and the supply side of the market, while confidential information pertains more often to customer information or the demand side of the market.

\textsuperscript{16} Van der Werf, Ziegeleara and Schoenmakers (1997).

\textsuperscript{17} ‘t Hart (1979); Curley (1994).
Chapter 5

It is less clear, however, that this solution is also efficient from society’s point of view. Once it has been produced, information is non-rival, which means that once it has been used or consumed by a producer or a consumer, the original amount is still available for others. Information is thus a non-scarce resource, and efficiency demands that it should be made available free of charge (apart from the cost of transmitting the information). On the other hand, if producers could not recoup the costs of their investments, they would have no incentive to produce valuable information. This is why intellectual property rights were created. By excluding others from the use of valuable information it becomes possible to appropriate much of their value. However, a covenant not to compete is fundamentally different from intellectual property rights. A property right to information excludes everybody but the owner from its use, while a covenant not to compete excludes only the other party to the contract: the employee. Competitors are still free to find other legal ways to extract the information or to develop and then use the information themselves. Nevertheless, by allowing the protection of information through noncompetition clauses, the loss to society in terms of higher producer costs for competitors or a decline in consumer utility remains. On the other hand, there is a gain to society in terms of an acceleration of the production of valuable information and the associated future decline in costs of production or increase in consumer utility. The optimum level of protection is where the marginal social cost of protection is equal to the marginal social benefit, which will depend on such variables as the duration of the noncompetition clause, the substitutes available to competitors and the likelihood of duplication in creative efforts by competitors.

5.3.3 Noncompetition clauses and self-selection

The arguments discussed so far do not seem to explain fully why noncompetition clauses are so often included in employment contracts. Nor do they explain

---

18 It has been suggested by Gilson that Silicon Valley was a success because noncompetition clauses were prohibited, see R.J. Gilson (1999), ‘The Legal Infrastructure of High Technology Industrial Districts, Silicon Valley, Route 128 and Covenants Not to Compete’, New York University Law Review, pp. 575-629.
why apparently only a small proportion of employees who are legally constrained by these clauses feels restricted by them\textsuperscript{19}. In this and the following section, I will analyse the economic theories of employment relationships for some of their implications on bargaining and contracting by the employer and the employees. The aim is to predict the circumstances under which noncompetition clauses are likely to arise.

For several reasons, employers may have an interest in attracting to the firm employees who are seeking long-term employment. These reasons may include the costs of employee turnover, the need to entrust the employee with confidential information or the employer’s wish for employees to develop personal relationships with clients, which usually takes time. If the firm succeeds, it will be more productive and more value will be created and available for distribution between the employer and the employees. However, if the firm were to offer long-term employment contracts including a surplus over the market wage, without additional guarantees, applicants could not credibly claim they prefer employment for an extended period of time. They could simply accept the higher paid job and resign when it suited them. The employer, of course, knows that some of the applicants would not be fully committed to the contract offered and could attach a noncompetition clause to it\textsuperscript{20}. Suppose now that there are two groups of employees, one preferring to stay longer and the other to stay for a shorter period of time with the firm. Both groups would be paid a higher than market wage. If an employee were to resign, the noncompetition clause would prohibit the employee from being employed in the same line of work for some time. This would be a costly restriction on freedom of choice, especially for those employees whose preference is to change jobs often. An employee might

\textsuperscript{19} More than 75\% of the employees who entered into a noncompetition agreement do not feel restricted by the the clauses. Of those employees who are looking for a new job, about one fifth feel restricted by the noncompetition agreement, see C. van der Werf, A. Ziegelaar, F. Schoenmakers, \textit{Het concurrentiebeding in de praktijk}, Research voor Beleid, Leiden, 1997, pp. 24-28.

\textsuperscript{20} In the literature, a number of other alternatives to induce an employee to stay with the firm has been analysed, including efficiency wages, promotion policies and pensions. See generally O. Ashenfelter and R. Layard (eds.) (1987), \textit{Handbook of Labor Economics}, Amsterdam, North Holland.
Chapter 5

expect the employer not to enforce the clause because of the time, energy and money it would cost him. To play down this hope, employers often strengthen the noncompetition clause with a liquidated damages clause. By offering employment contracts including a noncompetition clause, different groups of employees would effectively self-select for those jobs where long-term employment is desirable and where it is not.

One may question whether an employer would not do better by offering a contract for, say, five years of employment as an alternative to a noncompetition clause. But there are both legal and economic reasons why such a contract would be no better than a second-best alternative. From a legal point of view, both the employer and the employee would not prefer such a contract. From the point of view of the employer, such a contract would limit his freedom to fire the employee, and if the courts were to fail to enforce the contract, he would have to pay damages. For an employee, it would no longer be possible to resign at all times. Moreover, if the object of the contract was to establish a long-term employment relationship that would be favourable to induce investments from both sides, this end would not be achieved. This can be seen easily by looking at what would happen in the final year of the employment contract and then working backwards to figure out what both parties to the contract would do along the way. In the fifth year of the employment contract, neither party would invest if they knew it was the final year of their employment relationship. But if they were not going to invest in the fifth year, they would know this in the fourth year of their employment relationship and they would not invest in that year either etc. Therefore, the employment contract for a specified period of time is not a type of contract to develop a productive long-term employment relationship.

If noncompetition clauses were to self-select employees with a preference for long-term employment relationships, it would explain why so many employees do not seem to have any problem in agreeing with such clauses.

---

21 Generally, liquidated damages are enforceable if they reflect a reasonable estimate of the actual damages expected to arise from the breach of the contract.
5.3.4 Noncompetition clauses to prevent opportunistic behaviour and renegotiation

After completion of the process of search and selection, an employment contract will be concluded between the employer and the employee. Ideally, such a contract would encompass a description of all relevant aspects in all possible circumstances to both parties, a determination of a desirable course of action in each situation and compliance and enforcement mechanisms to determine if the terms of the contract have been met or violated and to enforce agreed performance. However, for reasons of bounded rationality, the cost of writing contracts and the cost of enforcing contracts, the typical employment contract is a very incomplete contract\(^\text{23}\). Instead of describing the content of the work very precisely, the employer is given the power to direct the employee’s actions. The employee agrees to put in his time and best efforts whereas the employer will pay him a wage. Not only is an employment contract incomplete, but much of what the parties expect and promise is also left implicit. All this saves on transaction costs. However, other transaction costs are likely to increase. If important aspects of the employment relationship have not been written down or have been left implicit, how can both parties be sure that their understandings will be honoured and their expectations will be fulfilled? The purpose of the employment relationship is the creation and distribution of a surplus of value over and above the value of the resources invested by parties involved. But if the terms of the contract have not been written down and cannot be verified, how can one be confident that parties are committed to their (implicit) promises and obligations. More specifically, a party may invest resources specific to the relationship, which cannot be withdrawn when the relationship ends. These investments make this party vulnerable to opportunistic behaviour by the other party. The other party may, for example, claim that circumstances have changed such that renegotiation of the terms of the contract becomes necessary or she may claim that the (unverifiable) terms of the contract are such that the division of the surplus should be different from what it currently is. This so-called hold-up problem may arise when an employer or an employee has incurred

substantial sunk costs in establishing and/or developing the employment relationship\textsuperscript{24}. Examples are an employee who buys a home near the location of a plant, or an employer who provides firm-specific ‘on-the-job’ training. When contracts are incomplete, parties to the contract may try to renegotiate better terms than they originally agreed to. Both parties will anticipate the possibility of opportunistic behaviour and their adverse effects on investments and productivity. Furthermore, both parties will also be aware of the costs of the bargaining process itself and the possibility that renegotiation may not lead to an agreement. Consequently, they will seek ways to prevent renegotiation of the contract and to achieve commitment to the agreed clauses. In the literature, a number of institutions have been analysed that may reduce opportunistic behaviour\textsuperscript{25}. In the context of employment contracts, fixed wage contracts can overcome the hold-up problem\textsuperscript{26}. Efficient investment is ensured if wages are fixed independently of those investments.

However, in a number of cases, fixing wages may be undesirable. To align the interests of the employees with his own, an employer generally has two options. He can either monitor the performance of the employees or he can devise incentive-pay schemes that link pay to performance. In some cases, monitoring may be prohibitively costly. One can think of those tasks where external contacts or close personal relationships with the clientele are essential\textsuperscript{27}. In other cases, it may be more obvious to link pay to performance, such as with salespersons or managers. It is no accident that noncompetition clauses are often found in contracts with regard to just those kinds of employment. There are large costs involved in fixing the wage, and employee performance is permanently deliberated. At the same time, the employer will have made himself vulnerable by supplying the employees with strategic information of the firm and


\textsuperscript{25} Dixit and Nalebuff (1991); Schelling (1998).

\textsuperscript{26} Malcomson (1997).

\textsuperscript{27} Another example is research and development activities, in which it is possible to monitor the time spent on the activity, but not the relevant mental dimensions of the activity.
of its clientele. In a process of deliberation and implicit bargaining these employees could easily capture some of the returns on the employer’s investments by threatening to leave the firm and use the information themselves. To prevent this kind of hold-up and to promote an employment relationship that is conducive to investment and effort from both sides, a noncompetition clause is included in the employment contract. On the one hand, the threat of leaving the firm is no longer credible and the employees need not worry whether they should renegotiate the terms of the contract or not. The outside options are now less favourable than staying with the firm. The employer need not fear the loss of valuable information. He can safely invest in the firm and make his valuable resources available to the employees. On the other hand, the employees need not fear a hold-up by the employer. The employer knows that the best efforts of his employees are essential to the creation and maximization of the surplus of the employment relationship. By linking the wage to their performance, employees are motivated to put forth their best efforts. Figure 1 summarizes the foregoing analysis.

Suppose that a worker would have to change residence or incur a large search cost for a new job. The worker now obtains a quasi-rent if she continues the employment relationship because her job is more valuable than a comparable job would be. The employer may have invested in firm-specific training of the employee or he may have incurred substantial recruiting costs. The employer also obtains a quasi-rent if he continues the employment relationship with the employee because the worker is now more valuable than a replacement worker would be. In figure 1, \( W_w \) is lowest wage the employee will accept before turning to an outside offer\(^28\). The highest wage the employer is prepared to pay is given by \( W_c \). The total surplus of the employment relationship, the distance \( (W_c - W_w) \), is determined by the specific investments and the level of turnover costs of the employer and the employee. In period 0, the contract is negotiated and the wage is determined such that the surplus of the relationship is divided equally between the employer and the employees: \( W_c \). Let us assume this wage is partly fixed and partly depends on performance, for example sales. In period 1, the employer makes certain trade secrets or customer details lists available to

---

\(^{28}\) Outside options may fluctuate, for instance because of the business cycle.
the employee. We assume that this investment of the employer is made at once and is immediately sunk. Together with advertising campaigns, provision of information to customers, the handling of sales by the firm etc, the employee is able to develop a close personal and productive relationship with the clients. The productivity of the employee increases in period 2. Her outside offer increases to, say $W_o$.

![Diagram](image)

**Figure 5.1   Hold-up and noncompetition clauses in employment contracts**

In the absence of a noncompetition clause the employee may now hold up the employer. The threat to leave the firm and to disclose the confidential and valuable information to a competitor is credible since it will be very difficult to prove that information has been revealed or damages were incurred. The most extreme result of executing the threat would be the capture of all the firm’s revenues and profits by the competitor. Renegotiation takes place in period 2.
The surplus of the employment relationship declines to the distance \( (W_e' - W_o) \). The current employer gives in and pays a higher wage. A rational employer would anticipate the possibility of ex post opportunistic behaviour by the employee and would not invest in the relationship in the first place. The situation would be completely different, however, if a noncompetition clause was included in the employment contract. Now the employee cannot credibly threaten to leave the firm because in that case his best outside offer would be lower than his current wage, let us say \( W_n \).\(^{29}\) For both the employer and the employee it would be best to agree to an employment contract that includes the noncompetition clause. If contracts are incomplete and the terms of the contract cannot be verified or enforced by the courts then noncompetition clauses may achieve commitment to the implicit or non-observable terms. The noncompetition clause is beneficial to both since they both obtain quasi-rents that they would lose if the employment relationship were to end. Figure 5.1 illustrates the effects of noncompetition clauses: they enhance the value of the employment relationship and increase the possibilities of efficient exchange between the employer and the employee.

5.3.5 Undue hardship and reputation

Legal scholars often complain that employers persistently and hard-headedly enforce their noncompensation agreements. This is one of the reasons why legislatures and courts frequently interfere in the freedom of contract on behalf of the employees. Not only are the terms of the contract unbalanced, but employers seem determined to stick to their contractually agreed rights. The theory laid out above explains why employers are so tenacious in maintaining the noncompetition agreements. Employers need to have a reputation for enforcing their noncompensation agreements. The reputation is a guarantee that contracts will be honoured. In the absence of this reputation, an employee would feel the temptation to renegotiate the terms of the contract once the employer has invested in the employment relationship. The employee may even feel that rene-

\(^{29}\) The outside offer \( W_n \) is also lower compared with the outside offer in the situation in which no noncompetition clause is included.
Chapter 5

gotiation is justifiable, given her special efforts and her special relationship with clients, which makes the employment relationship so productive. But employers would not invest in the employment relationship in the first place if they knew that they could not recoup the costs of these investments. Depending on how important noncompetition clauses are for other employment contracts of the firm, the importance of long term employment relationships and the quasirents that are being obtained, one may predict that employers will build a reputation in enforcing noncompetition clauses.

The theory also suggests that courts should adopt a different perspective in determining the reasonableness of noncompetition clauses. Commonly, the reasonableness of these clauses is determined by comparing ex post the interests of the employer, the employees and the public. However, because noncompetition clauses increase the value of the employment relationship, it seems likely that employees will be compensated for the inclusion of these clauses in their employment contracts. If the aim is to determine whether or not noncompetition clauses are reasonable, courts should compare the ex ante present value of employment contracts with and without noncompetition clauses. Such a comparison would make clear whether or not employees are compensated effectively for the inclusion the noncompetition clauses and whether these clauses are balanced or not. If the courts, with their ex post comparison of the interests of the employer and the employees, were only to bring about a transfer between employers and employees, not much harm would be done. However, by making this comparison and by potentially undoing an agreed contract between an employer and an employee, courts effectively decrease the ex ante value of the employment relationship. The result will be lower investments and lower wages.

5.5 Summary and conclusions

Generally, legislators and judges are critical of noncompetition agreements. The terms of such contracts are often deemed to be unreasonable. According to the underlying legal theory, the unbalanced terms find their origin in an unequal bargaining position of the employees, and in transaction costs and informa-
tional asymmetries. The economic analysis questions the assumptions underlying the legal theory on noncompetition clauses. Why shouldn’t employers be effectively disciplined by market forces? Furthermore, if noncompetition clauses are costly for the employees, they will spend their resources in searching for employment contracts without those clauses or to be adequately compensated if these clauses are to be included in the employment contract. Moreover, given the type of employment where noncompetition agreements are customary, it is unlikely that employees will be unaware of the terms and of their consequences. These employees may be expected to have had sufficient education or training, are often experienced in their jobs, and it is in their interest to know all about the noncompetition clauses. The question then arises as to why employees sign contracts that seem so unreasonable and unbalanced. Four theories are advanced in explaining noncompetition agreements. In the literature, the argument is often made that noncompetition clauses ensure that the employer will have his investments in ‘on-the-job’ training refunded in the event that the employee leaves the firm shortly after being trained. An analysis of this argument makes clear that both employers and employees will generally prefer pay-back clauses to noncompetition clauses in their employment contracts. Another theory explains noncompetition clauses from the desire to protect trade secrets and confidential information. Although there are alternative legal options such as a tort action or covenants not to disclose, it is difficult to prove that a trade secret or confidential information has been revealed, and it is often even more difficult to prove damages. Depending on the costs of compensating employees, noncompetition clauses may be an efficient solution to the potential loss of valuable information. The third theory explains noncompetition clauses from self-selection. Employers do not possess information on the true characteristics of potential employees. For some types of employment, they want to enter into long term employment contracts and are willing to pay higher wages. Noncompetition clauses are used as a device for self-selecting employees who desire a long term contract with the firm. The fourth theory explains covenants not to compete as an instrument to prevent renegotiation and hold-up in the employment relationship. It is most likely that both parties to the employment contract are well aware of the noncompetition clauses and their consequences. They agree to these terms to guarantee that specific investments will be made in the employment relationship. The noncompetition agreement
Chapter 5

protects the creation and distribution of the surplus of the employment relationship.

Acknowledgement
The author is grateful for the valuable comments on an earlier version of this paper made by Prof. Dr. Gerrit De Geest, Prof. Dr. Roger Van den Bergh and Prof. Dr. Jacques Siegers, none of whom may be held responsible for the views expressed in this paper.
CHAPTER 6 SUMMARY AND CONCLUSIONS

The final chapter of this volume starts by summarizing the findings as they relate to the particular fields of the research. The main results of the preceding chapters are then brought together and a number of overall conclusions are drawn. The chapter concludes with a statement of policy implications and plans for future research.

6.1 Summary and conclusions of the particular fields of research

6.1.1 General Theories of Regulation

Chapter one presents a review and synthesis of the economic literature on regulation. It is customary to make a distinction between public interest theories and private interest theories of regulation. Public interest theories explain regulation from the pursuit of the public interest, which may be defined as the efficient allocation of scarce resources for individual and public goods. Public interest theories take market failures as their point of departure and assume not only that the state is the most efficient institution to deal with market failures but also that it is willing to pursue the goal of the efficient allocation of scarce resources. Market failures may result from imperfect competition, unstable markets, missing markets and undesirable market outcomes. The regulations are intended to neutralize market failures and maximize welfare.

The literature raises several criticisms of these theories. What are referred to as ‘market failures’ may be described more satisfactorily as ‘model failures’ because the analysis takes no account of transaction costs. Society is assumed to behave as a perfectly competitive market economy with no transaction costs and it is only through the lens of this model that market failures appear. If an appropriate economic model was available that included transaction costs and informational problems, all behaviour would be qualified as efficient and market failures would disappear. The conclusion that regulations would improve efficiency could only be drawn after an analysis of alternative institutions allo-
Chapter 6

cating resources. The assumption that the state is the most efficient institution to deal with market failures has been criticized in both empirical and theoretical research. One of the theoretical arguments is that it is precisely the transaction costs and information costs that cause markets to fail and that are assumed to be absent in government regulation. From an empirical point of view, several studies have pointed to inefficiencies and the ineffectiveness of specific regulations and to efficiencies that arise from the deregulation of certain industries. Although these criticisms have led to transaction costs and information costs being incorporated into the public interest explanation of regulations, a number of criticisms remain. Most importantly, these criticisms have to do with the usual narrow definition of public interest that equates it to economic efficiency. However, policy aims other than economic efficiency such as procedural fairness or distributional justice often figure more prominently in politics. If a trade-off is involved between economic efficiency and these other objectives, the public interest theories are unable to explain adequately the type and kind of regulations we find in practice. If, on the other hand, the public interest is defined more broadly as social efficiency, practical difficulties for testing and refutation arise. Not only will it be difficult to operationalize social efficiency, but its meaning will most certainly change in the course of time. Another criticism is one of incompleteness in that, in contrast to the theory of a perfectly competitive market economy, the public interest theories of regulation fail to indicate how a given view of the public interest translates into legislative actions that maximize efficiency. Finally much evidence remains that would appear to refute the public interest theories of regulation.

With a view to counterbalancing the public interest theory, an alternative explanation of regulation was expounded from the so-called private interest theories of regulation, in which the purpose of regulation is to transfer income or wealth in exchange for political support. In a seminal paper, Stigler maintained that ‘as a rule, regulation is acquired by the industry and designed and operated primarily for its benefit’. An industry may benefit from regulations such as minimum prices, entry or exit barriers, minimum quality standards and such like. These regulations serve to limit competition and facilitate transfers from consumers towards producers and among producer groups. Transaction costs and organization costs mean that consumers will find it difficult to organize and
trade in the political sector, which is why regulations will generally favour small producer groups at the cost of large consumer groups. In an extension of this theory, the core problem for regulators was defined by Peltzman as efficient regulation. Regulators will not protect producers if a small departure from a profit-maximizing price level offers more votes from some consumers than the resulting loss of votes from the producer groups. This extension explains the cross-subsidization of specific consumer groups and the regulation of both competitive and monopolistic industries. Although the transfers associated with regulation bring about welfare losses, the competition among pressure groups, according to Becker, will set a limit on these welfare losses. The higher the welfare losses, the more countervailing pressure will develop and the less likely politicians will be to put such regulations into effect.

Several criticisms have been raised against these theories. Firstly, regulation is seen as the outcome of a bargain in which transfers and votes are traded. However, regulation will generally have redistributive effects and if these are now seen as the main driving force of regulation, the theory risks becoming a tautological. Secondly, the theories expounded so far are unable to predict which groups will be the most effective politically and which groups will be favoured with transfers. Originally, the presumption was that producer groups would benefit most from regulation, but empirical research has revealed that it was the workers who actually did so. Thirdly, the causal mechanisms by which politically effective groups are able to control the legislature are strikingly absent, as is the legislature or the administration. These theories generally lack an explanatory analysis of the behaviour of political actors such as voters, members of parliament, legislators, bureaucrats or agencies and courts, individually or collectively in the regulatory process. A separate field of research, known as public choice theory has developed to deal with these matters. The private interest literature of regulation includes several attempts to bridge the above mentioned gaps, but a general private interest theory of regulation has yet to emerge.
Chapter 6

6.1.2 The Explanation of Deregulation in the Netherlands

In the Netherlands, deregulation has been a political priority since the 1980s, but actual results were not forthcoming until the mid 1990s. This paper has applied theories of regulation to derive hypotheses on the possible causes of deregulation. A distinction is made between public interest theories of regulation, the Chicago theories of regulation and public choice theories of regulation. The Chicago theories are both the foundation and a constituent part of the private interest theories of regulation. The fundamental building blocks of the public interest theories are market failures and the government as the most efficient institution to correct market failures. Two major driving forces behind deregulation may be predicted from the public interest theories of regulation. Firstly, demand or technological developments may have eliminated the market failure and secondly, regulation may no longer turn out to be the most efficient way of correcting the market failures. The Chicago theories of regulation suggests are four circumstances under which deregulation is likely to occur. Firstly, changes in the cost of information and organization may turn disadvantaged parties of regulation into a politically effective group. Secondly, the current politically effective groups may believe that they can advance their economic interests better in an unregulated market, possibly by way of self-regulation. Thirdly, a decline in the industry’s profits, for example as a result of lower demand or higher input prices, may upset the political balance between the opposing pressures. Fourthly, if deadweight losses increase over time, external pressures may undermine the bias in favour of the status quo.

The public choice theories offer the following explanations of deregulation. Firstly, there could be a change in the balance of interest groups. Secondly, political entrepreneurs, such as the chairmen of the regulatory commissions, could disrupt the status quo. Thirdly, politicians may successfully seek political support for deregulation by providing the electorate with information on the inefficiencies of regulation. Alternatively, politicians could try to use the complexi-

---

1 The public choice theories of regulation, unlike the public and private interest theories of regulation, analyse explicitly the behaviour of actors in the regulatory decision-making process and their interaction.
ties of regulatory issues by claiming that deregulation would greatly advance economic and social welfare. After an analysis of the organization, the circumstances and the results of deregulation in the Netherlands, it is concluded, that of the three groups of theories, the public choice theories seem to have the greatest explanatory power of deregulation in the Netherlands. Two variables have been identified that have made a particular contribution to the effectiveness of the policy of deregulation: public opinion and political entrepreneurship. The first variable changed the balance of power of political parties. After the 1994-elections, the Christian Democrat Party was excluded from the government coalition for the first time since 1918. The second variable, political entrepreneurship, was crucial in revising competition policy and in other deregulatory proposals. The former chairman of the Committee on Deregulation was appointed secretary-general of the Department of Economic Affairs, which put him in a position to pursue his goals of effectively controlling the bureaucracy and furthering the policy of deregulation.

6.1.3 An Economic Analysis of (Self-)Regulation of Physicians in the Netherlands

This paper seeks to find an explanation for the regulation and self-regulation of physicians in the Netherlands. There are severe restrictions on competition in the market for medical services. Only a limited number of students are allowed to enter medical school, where the course takes a formidable twelve years, which is about three times as long as the average academic study. After six years of study, students have to specialize as a general practitioner, a medical specialist or a social physician, such as a school doctor. Unlike other European countries, the education, official recognition and registration of these specialisms is organized and performed entirely by the profession itself. Only a physician is entitled to perform any medical interventions and the title of physician is reserved by law for graduates of medicine who have taken the Hippocratic oath. By specializing, a physician voluntarily agrees to restrict his domain to certain specialized medical services. If a physician steps outside his domain he risks being prosecuted under statutory disciplinary law. The profession, the
health insurance institutions and the hospitals have also set up a number of other restrictions to entry. For instance, there are fixed prices for all medical interventions, which are determined in negotiations between the physicians’ organizations and the insurers. Once approved by a public institution, no new or higher prices may be charged without its approval. Furthermore, rules of conduct stipulate that doctors may not entice patients away from colleagues, nor advertise professional skills or specialized methods of treatment. Finally, when collaborating with other doctors, they are not allowed to choose the organizational form they see fit: hierarchical structures among doctors or associations of different specialisms, are deemed undesirable. Quality regulations pertain mostly to the structure and procedures of the medical services rendered, not to the outcomes of the services themselves.

From a public interest perspective, a number of explanations have been offered to justify these regulations. Medical services may be characterized as experience goods or even trust goods. With free entry and exit, quacks and charlatans could enter the market and patients would not know if a lower price implied a cheaper service or signalled lower quality. Consequently, a process of adverse selection could develop whereby high quality services are driven out of the market by low quality services. A higher education organized by the profession and fixed prices would guarantee minimum quality standards. Sickness is an uncertain event and adverse selection is a well-known phenomenon in insurance markets. Compulsory insurance has facilitated access to health care, but it may also result in unrestricted demand for medical services. On the supply side, physicians acting in accordance with professional standards would offer these services. For this reason, and to prevent excess demand created by the supply of medical services, a limit should be set on entry into the market for medical services.

A number of other empirical facts seem to refute the public interest theory of regulation of the medical profession. Several unexplainable facts have been documented. For instance, even if corrected for the structure of the population and socio-economic variables, there are considerable differences in the production of medical services from one doctor to another. This suggests large differences in quality among doctors despite the quality regulations. Furthermore, for some diagnostic medical services, there is evidence that a high proportion of the procedures performed are superfluous. Another fact that seems to be in con-
flict with the public interest theory is the level of income of the medical specialists. Although considerable differences exist between specializations, these doctors enjoy earned income in the top 1% in the Netherlands, and they account for a large proportion of this group.

An alternative explanation for these facts is suggested by the private interest theories of regulation. According to these theories, regulation is an instrument for bringing about hidden transfers in exchange for votes or to prevent the loss of votes. A number of facts are consistent with these theories. Medical specialists form a small group of only 7500 doctors, about 80% of whom belong to the representative organization. The cost of organization is low and most of these doctors already belonged to academic communities, whereas the cost of organizing the constantly changing group of numerous patients would be comparatively high. Still, the information on medical services and their costs is asymmetrically distributed. What is more, the marginal deadweight loss of redistribution would probably be small whereas doctors benefit significantly from a small increase in fees. Becker’s theory of countervailing pressure is valid. The increase in the macroeconomic cost of the health care sector and the associated increase in insurance premiums and labour costs endangers employment and economic growth. This presents a threat to the re-election of politicians, who consequently curtail this increase in health care outlays.

### 6.1.4 Reversed Solidarity in Pension Plans

Chapter five applied the economic literature on regulation and public choice to explain reversed solidarity in pension plans. Pension plans are a mixture of regulation and individual and collective bargaining. Reversed solidarity is a consequence of defined-benefit pensions funded on the capitalization principle, where the pension benefits depend on final earnings and contributions are specified as a uniform percentage of income. Most pension plans in the Netherlands have an element of reversed solidarity, in which people with no career path contribute to the pension provisions for those with a career. In general, reversed solidarity comes down to blue-collar workers and women paying for the pensions of white-collar workers, most of whom are men. This dimension of pension plans is surprising in view of the fact that their funding is regulated by
law and unions occupy half the seats on the pension fund boards. It is usually assumed that unions alter pension plans in ways that equalize pensions among workers. Why do workers not bargain for a cheaper or more rewarding pension contract and what is the explanation for reversed solidarity in those contracts? The analysis makes clear that information and transaction costs are prohibitive to concluding better pension agreements. Firstly, the law on pensions stipulates ‘automatic’ participation for a new employee entering a group with an existing pension agreement. The employer is legally obliged to deduct the necessary pension premiums from the wage. Secondly, the termination of the pension agreement usually coincides with that of the labour contract and the pension funds do not supply a copy of the standard form pension contract until the employee has joined the firm. Thirdly, weighing up alternative pension contracts would entail a formidable task of comparing, contrasting and evaluating many different types of contracts. Fourthly, employees would feel reassured about the terms of the pension plans in view of the fact that management and the higher paid workers were also covered by the same pension plan.

Pensions are the outcome of a bargain between the union and the employers’ organization. Employers may favour pension plans with reversed solidarity characteristics. One advantage is that withholding some of the employees' wages in the form of pension benefits until completion of their employment encourages them to continue to perform. Furthermore, a compensation scheme with low wages at the beginning of the career and increasing wage rates with length of tenure makes the firm attractive for low mobility and high productivity employees. The pension plan can therefore be seen as a screening device to identify such workers for senior posts. A pension plan based on final earnings increases the effectiveness of this screening device because of the associated costs of leaving the firm. When a number of employers participate in an industry pension fund, the Minister of Social Affairs may declare participation compulsory. In the Netherlands, the majority of pension plans are executed by industry pension funds, in more than eighty percent of which participation is compulsory. This may help explain why unions are prepared to overlook reversed solidarity in pension plans. The fact is that when employers opt for an industry pension fund, statutory law dictates that unions are entitled to occupy half the seats on the board, enlarging their control over the investments of contributions, even from non-union members. It may also further the aims of union
leaders, because not only does it endow them with more prestige and influence, it is also likely to increase union membership. Furthermore, as a collective democratic organization, the union will represent the preferences of the median voter rather than of the marginal employee. Since total compensation increases with seniority, even for non-career employees, median voters who are older and closer to retirement, would prefer defined benefit plans. Finally, uniform pension premiums throughout the industry are analogous to the ubiquitous standard rate wage policy of the unions. A comparison of the pension plans of Germany, France, Belgium and the United Kingdom reveals that some of the pension plans in Belgium and the UK also exhibit reversed solidarity.

6.1.5 Noncompetition Clauses: Unreasonable or Efficient?

One in every five employees in the Netherlands enters into a noncompetition agreement. A noncompetition agreement, or ‘covenant not to compete’, protects the employer from competition by the employee after his employment has ended. The agreement may prohibit the employee from working for a competitor in the same region, or for an employer in the same sector, product market, or line of work, for a certain period of time after the employment has ended. In many countries, legislatures and courts have felt the need to protect employees against these clauses. Some places, such as California and Mexico, simply forbid these contractual restraints. Elsewhere, such as Belgium, Germany and Spain, noncompetition clauses have to comply with certain standards laid down in law in order to be valid. In the Netherlands, the law specifies only a few formal conditions for noncompetition agreements: they have to be written and agreed by employees who are above the age of majority. However, an amendment of the law that is in preparation, will lay down stringent restrictions on the use of these clauses. Generally, legislators and courts tend to disapprove of noncompetition agreements, judging them to be unreasonable and rooted in the unequal bargaining position of employees, and in transaction costs and informational asymmetries. My paper questions the assumptions underlying the traditional legal theory on noncompetition clauses. Employees also agree to noncompetition clauses in competitive labour markets where they do not have an unequal bargaining po-
position. Furthermore, given the type of employment for which noncompetition agreements are pertinent, such as management or research and development, it is unlikely that employees are unaware of the terms or of their consequences. Four theories are advanced to explain noncompetition clauses. The literature often mentions the argument that noncompetition clauses ensure that the employer will be able to recoup his investment in ‘on-the-job’ training should the employee leave the firm shortly afterwards. An analysis of this argument makes it clear that both employers and employees will generally prefer payback clauses to noncompetition clauses in their employment contracts. Another theory explains noncompetition clauses from the desire to protect trade secrets and confidential information. Although alternative legal options such as a tort action or non-disclosure covenants exist, it is difficult to prove that a trade secret or confidential information has been revealed and often even more difficult to prove damages. Depending on the costs of compensating employees, noncompetition clauses may be an efficient solution to the potential loss of valuable information. The third and fourth theories have not previously been encountered in the legal or economic literature on noncompetition agreements. The third theory explains noncompetition clauses from self-selection. Employers do not possess information on the true characteristics of potential employees. For some types of employment, they want to enter into long-term employment contracts and are willing to pay higher wages. Noncompetition clauses are used as a device to self-select employees who desire a long-term contract with the firm. The fourth theory explains noncompetition covenants as an instrument for preventing renegotiation and hold-up in the employment relationship. It is most likely that both parties to the employment contract are well aware of the noncompetition clauses and their consequences. They agree to these terms to guarantee that specific investments will be made in the employment relationship. The noncompetition agreement protects the creation and distribution of the surplus of the employment relationship. Legislatures and courts interfere in noncompetition agreements not only because the terms of the contract appear to be unbalanced, but also because employers seem so tenacious in maintaining these agreements. The analysis makes it clear that employers need to build a reputation in enforcing noncompetition clauses. The theory suggests that courts should adopt a different perspective in determining the reasonableness of noncompetition clauses. Instead of comparing *ex*
post the interests of the employer and the employees in such cases, the courts should compare the ex ante present value of the employment contract with and without noncompetition clauses. By making an ex post comparison, courts undo an agreement between an employer and an employee and by doing so they effectively decrease the ex ante value of the employment relationship. The result will be lower investments and lower wages.

6.2 Overall conclusions on public and private interest theories of regulation

The literature reveals a debate on the explanatory power and the scientific status of the theories of regulation, with the public interest theories being called into question most frequently. Appraisals of public and private interest theories of regulation require criteria with which to judge these theories. The criteria may be inferred from the methodology that is regarded appropriate when conducting economic research. The typical methodology of economics in explaining a certain problem or event is to discover and formulate a theory and then to deduce the event from the theory, given a number of initial conditions, auxiliary assumptions and boundary conditions. The theory may then be used to deduce predictions of particular events about which nothing is yet known. The theory is tested by inspecting whether the predictions are borne out by observation. The theory is confirmed or disconfirmed depending on whether the predictions turn out to be accurate or inaccurate. For example, to explain the regulation of the power industry, first it is observed that prices are not equal to marginal costs and it is concluded that the market for power operates imperfectly. Then, a number of assumptions are made, for example with respect to transaction costs and information imperfections, and it is concluded that the government will regulate this market to improve economic efficiency. This

---

2 See for example Kaserman and Mayo (1995, p. 518) and Viscusi, Vernon and Harrington (1996, p. 326), who even maintain that the public interest theories have lacked supporters for several decades.

3 For an introduction to the methodology of economics, see, for example, Blaug (1992) or Stewart (1979).
process creates a ‘market failure theory of government intervention’, which predicts that regulation exists in other markets or sectors where prices are unequal to marginal costs, or where other forms of market failures are present. A test is performed to determine whether these markets exhibit government intervention and thus whether the market failure theory of government intervention is upheld. The testing of a hypothesis requires initial assumptions and a number of supplementary hypotheses. If the prediction has been disproved it may still be possible that a change has taken place in the initial assumptions or that some of the supplementary hypotheses are false. The rejection or acceptance of a theory thus rests not only on empirical evidence but also on the plausibility of its initial assumptions and supplementary hypotheses. This mainstream methodology of economic research makes clear that different criteria will be applied to appraise and justify the different theories of regulation. Examples of such criteria are: the degree of testability, the degree of confirmation or corroboration, logical consistency, rigorous formulation or elegant construction, whether it proceeds from the principles of methodological individualism, the completeness of the theory, coherence with other accepted theories, whether its assumptions are realistic, whether it stimulates further research, its practical relevance, and so forth. If such criteria are applied to the public and private interest theories of regulation, is it possible to accept or reject either of them unambiguously?

6.2.1 Comments on the public interest theories of regulation

To be able to test the public interest theories of regulation, the concept of the public interest has to be clear and be given empirical meaning, but, unfortunately, the literature contains divergent definitions of ‘the public interest’, and it is even more difficult to give empirical content to these definitions. For example, it is customary among economists to define the public interest as the efficient allocation of resources, which describes a situation where resources such as labour, capital and natural resources are deployed to their highest valued uses. But there are different interpretations of these ‘highest valued uses’. For an economist, it is natural to judge an allocation of resources to be efficient if prices of goods and services are equal to their resource costs (economic efficiency). Oth-
Summary and Conclusions

ers, however, have advocated a broader interpretation of efficiency that explicitly incorporates distributional, social or democratic values (social efficiency)\(^4\). The consequence of these different interpretations of the public interest is that the testing of the public interest theory of regulation may produce ambiguous results. For example, the regulation of the power industry may be interpreted as a refutation of the public interest theory, because prices have proven to be higher than marginal costs, but it may also be seen as a confirmation of the public interest theory in that these higher prices have made it possible to supply power to remote and sparsely populated areas\(^5\).

Another limitation of the public interest theories is that they do not include hypotheses on how market failure translates into a manifestation of the public interest and how this public interest then translates into political or regulatory decision-making to further the public interest. More specifically, the theories do not make clear how, in the event of a discrepancy between marginal social benefits and marginal social costs, uncertain benefits or individually non-exclusive but small benefits become aggregated such that a movement towards the equalization of marginal social costs and benefits sets in\(^6\).

Furthermore, although relatively little is known about the benefits and costs of most regulations, there is an increasing amount of evidence that seems to be inconsistent with the public interest theories of regulation\(^7\).

Finally, a number of questions remain unanswered. Firstly, for example, if deregulation is efficient, how come it took so long before it was accomplished and what is the explanation for the continuing growth of social regulation despite its apparent imperfections?

\(^4\) This broader interpretation of efficiency is the maximization of a specific social welfare function that explicitly ranks alternative allocations.

\(^5\) This ambiguity is one of the reasons why, according to Peltzman (1989), an economist only needs ten minutes to rationalize government intervention by constructing a market failure. It makes the public interest theory hard to falsify.

\(^6\) Furthermore, Arrow’s well-known Impossibility Theorem suggests that it may even be impossible to design institutions to aggregate the diverse preferences of individuals in a manner consistent with a set of reasonable conditions, see Arrow (1963).

\(^7\) Hahn (1998); Winston (1993). Hahn and Hird (1991) and Viscusi (1996), for example, show that a reallocation of resources for social regulations could save many lives at no additional resource costs.
Secondly, there are several solutions in theory to the problem of natural monopoly. However, the public interest theory does not make clear why the traditional solution in Europe has been to erect public enterprises while in the United States regulatory agencies were established to control and regulate private enterprises. Nor is it clear why at other times options such as franchise bidding or the introduction of competition were chosen. What is the explanation for the choice of standard setting as the traditional instrument in environmental and occupational health and safety regulation, while the efficiency of this particular instrument is broadly disputed and other instruments are advocated? What are the driving forces behind the recent establishment of new regulatory agencies and the introduction of market-like mechanisms such as tradable emission rights, auctions or standard-setting by private parties?

Thirdly, the public interest explanation of regulation partly explains regulation from the pursuit of distributional justice. Some examples of such regulations are: fair rate of return regulation, universal service requirements in public utility regulation, rent control, price floors for crops and minimum wages, affirmative action regulation, the use of subsidies versus taxes in environmental regulation etc. It should be possible to infer from the public interest theory of regulation who will be the beneficiaries of regulation and who will carry the costs. However, the public interest theories do not provide answers to this question.

### 6.2.2 Comments on the private interest theories of regulation

According to the private interest theories of government regulation, regulation is demanded by interest groups and supplied through the political process. Regulation is not established to correct market imperfections but to maximize and transfer rents and other benefits to politically effective groups.

Private interest theories assume the existence of interest groups, but theoretically it is still unclear how interest groups become organized, which strategies they pursue to influence regulators and legislators and which variables determine their effectiveness. Of course, a number of strategic variables have been identified, but the relationship between these variables and the regulatory out-
comes remains unclear\textsuperscript{8}. One example is the ambiguous relationship between the stakes held by different interest groups and the regulatory gains they stand to achieve. Besides that, a group of individuals with identical interests may be costly to organize, but may nevertheless be politically effective. Furthermore, it is uncertain whether interest groups attempt to have their preferred regulators elected or to alter the regulatory decisions through the currently elected regulators. These limitations make it difficult to test the private interest theories of regulation.

Furthermore, an adequate explanation of regulations demands a more complete analysis of the political system, including an explanation of the behaviour of legislators, regulators, bureaucrats, the judiciary, countervailing interest groups and voters\textsuperscript{9}. In order to determine whether private interests are the major determinant of regulations, we need to assess their relative influence in the political system, as opposed to the public interests, or the interests of the constituency at large. However, for practical and theoretical reasons, these interests are hard to define, model and measure.

From an empirical point of view, to explain particular regulations from private interest behaviour, in an ideal world one would like to relate regulatory decisions by politicians to activities by private interest groups\textsuperscript{10}. However, it is often difficult to directly observe the activities of private interest groups, such as influencing the electorate, lobbying among politicians, bribing individual legislators or financing political parties. Instead, the methodology followed in most economics research that sets out to explain regulations is to relate regulatory variables to the structural characteristics of the interest groups. This is a very indirect way of explaining regulations from private interest behaviour, and the results of this procedure may fail to give a clear-cut interpretation\textsuperscript{11}. Generally, private interest theories also give insufficient attention to market failures as driving forces of regulatory activities\textsuperscript{12}. There is also a significant amount of

\textsuperscript{8} On the formation of groups, see the classic contribution by Olson (1965).
\textsuperscript{9} Levine and Forrence (1990).
\textsuperscript{10} Potters and Sloof (1996).
\textsuperscript{11} For the same reason, private interest theories have been said to run the risk of being tautological (Noll, 1989).
evidence that seems to be inconsistent with the private interest theories of regu-
lration. For example, there are producer interest groups for virtually every eco-
omic sector, and consumers have significant disadvantages in information and
organization, yet we observe only a small number of rent-extracting price and
entry regulations in practice. The growth in social regulation seems largely to
be inconsistent with the private interest theories of regulation. The benefits of
these regulations are often dispersed among many individuals, as can be seen
in environmental regulation, while the costs are concentrated on relatively few
firms. Moreover, even if these regulations create entry barriers, protect incum-
bent producers from competition and extract rents from consumer groups, the
question remains as to why such costly instruments have been chosen to extract
and protect rents.
Finally, a number of questions remain unanswered. The private interest theo-
ries lack a clear explanation of recent regulatory developments, such as the use
of more market-like mechanisms, an increase in regulatory control and parlia-
mentary oversight, an emphasis on transparency and accountability etc.. These
developments put a limit on the creation and transfer of rents in exchange for
political support. Furthermore, what is the explanation for the choice of seem-
ingsly inefficient instruments such as standards instead of marketable emission
permits to acquire political support? And what is the explanation of deregula-
tion if profits and wages decline and consumers benefit?

6.2.3 General findings of the research

Against this background of conjecture and refutation on the explanatory theo-
ries of regulation, a first conclusion is that an unambiguous ranking of either the
public interest or the private interest theories of regulation in terms of being the
best in explaining regulations is (as yet) impossible\textsuperscript{13}. Much theoretical and em-
pirical work still needs to be done to understand how and why regulation is
chosen as the preferred instrument and why it takes one particular legal form
or another.

\textsuperscript{13} On conjecture and refutation in science, see Popper (1963).
A second conclusion is that the strength of public and private interest theories of regulation lies in their use as juxtaposing and opposing frameworks of economic analysis rather than in being corroborated theories. Public and private interest theories of regulation may be viewed as instruments for building up a model of the actual world and as instruments for understanding practical problems. These theories of regulation may be used as benchmark models or analytical frameworks to hypothesize on longer-term developments in regulations, on the political activities of interest groups, or on particular regulatory phenomena such as supernormal profits in certain regulated professions. The analytical frameworks thus stimulate research on the theoretical development of regulatory theories. Furthermore, the combined use of both the public and private interest theories of regulation will make it possible to discover new hypotheses concerning, for example, the circumstances under which one or another theory is likely to prevail. The use of both frameworks of analysis will also be helpful in developing and discovering an integrated theory of regulation that includes both private and public interests.

A third conclusion from my research on theories and practices of regulation pertains not to the usefulness of these theories for new theoretical developments but to the applicability of these theories to practical problems. The public and private interest theories of regulation may be interpreted as a box of tools for understanding and handling practical problems. For example, the use of both groups of theories made it possible to gain a perspective on such an unexpected result as reversed solidarity in pension plans, first by identifying possible market failures and then by applying private interest theories in the explanation. Furthermore, these theories can also be fruitfully applied to gaining an understanding

---

14 In the terminology of Klant, public and private interest theories of regulation can be seen as basic theories from which specific economic models are generated which are refutable, see Klant (1994, pp. 43, 67).

15 All these functions belong to the “context of discovery” rather than the “context of justification”. In the context of justification, knowledge is justified as scientific when certain criteria such as logical consistency and empirical validity have been met. The discovery of hypotheses of human behaviour, on the other hand, belongs to the “context of discovery”, see Blaug (1992, p. 15) and Popper (1959, pp. 31-32).
of certain observations and regularities. For example, one of the most common reasons for legal scholars to urge the regulation of noncompetition agreements is the tenacity with which employers enforce them. However, the economic analysis applied to noncompetition agreements has made it clear that such behaviour on the part of employers may be attributable more to the fear of being held-up by employees, than to an unequal bargaining position of the employees. More generally, the public interest theories can be applied to identifying possible causes of market failures and to summarize possible regulatory solutions, while private interest theories of regulation are applied to discovering where proposed or existing regulations are vulnerable to rent-seeking behaviour and may consequently display possible unexpected side effects or inefficiencies.

6.3 Policy implications

What is the relevance of the foregoing analysis for policymakers? To answer this question, a distinction must be made between the objectives, the instruments and the implementation of policy. Furthermore, the foregoing analysis has relevance for different levels of policymaking. First of all, the analysis has shown that the political participation of interest groups is likely to depend on the benefits of participation and also on the costs. If the objective of politicians is to pursue the public interest, they should minimize the information cost and transaction cost of participation. In general terms, this implies a wide dissemination of information on regulatory issues, open decision-making procedures and easy access to oversight committees, the regulators and the courts. Depending on the comparative efficiency of decision-making procedures, it may also imply covering the organization costs of relevant interest groups, such as consumers, pensioners and patients.

Secondly, since it is particularly the costs and benefits of regulation and regulatory decision-making that drive participation, the information about the probable effects of proposed regulation and the distribution of costs and benefits should be made widely available.
Thirdly, in the design of regulatory institutions, policymakers should be aware of the bias of participation in legislation and legislative decision-making. A decentralization of decision-making power is often an instrument against this bias and against potential abuse of regulatory decision-making by organized interest\textsuperscript{16}. However, there is a limit to the benefits obtained in this decentralization. If decentralized regulatory powers are extended to a case-by-case decision making power, information costs rise and the potential abuse by dominating interests increases. All other things being equal, this situation implies a preference for general rule-making rather than a case-by-case approach in regulatory decision making\textsuperscript{17}.

In the Netherlands, decentralization often takes place by delegating rule-making powers to ministers as heads of the departments. Until now, the preparation of legislation in the ministerial departments has been primarily the domain of legal scholars. Society stands to gain if legal scholars and regulatory economists intensify their cooperation, particularly when preparing legislation or regulation for decision-making. The standard monopoly assumptions often made by legal scholars do not seem to be consistent with the observations of market structures made by economists. If regulations are implemented to protect and improve the position of employees, these regulations may instead have adverse effects. More generally, normative legal analysis could gain from the use of economic analysis by founding its recommendations on a more thorough and complete observation of how markets operate. Economic science could also be useful in the analysis of the comparative efficiency of alternative legal instruments for pursuing social objectives such as distributive justice or equal opportunity.

\textsuperscript{16} Reliance on market-like instruments is another way to guard against dominating interests.

\textsuperscript{17} If, however, individual cases are characterized by a large degree of heterogeneity then the costs of general rule-making may outweigh benefits in terms of protection against private interests. Furthermore, it may be necessary in a period of transition from monopoly to competition to stimulate competition through asymmetric regulation, for example by imposing universal service requirements only on the former monopolist. Asymmetric regulation may require specific rule-making.
Chapter 6

A fifth implication concerns policymaking in the field of science. Legal and economic science could benefit from enhanced cooperation. The explanatory and predictive power of economic theories of regulation could be improved. For example, in the regulation of the network industries, economic surpluses are distributed between the suppliers of production factors such as owners and employees, and between or among consumer groups. A better insight into the trade-offs made by legislators and regulators in this distribution would improve economic models and enhance the predictive power of the nature of regulation and of the outcomes of the regulatory processes. Other trade-offs, such as between equality of treatment and the costs of legal procedures, also provide insights into the relevant legal values that furnish regulatory institutions and procedures. In the context of social regulation, for example, a better understanding would develop of the trade-offs involved in the legal choice of instruments and exceptions, or the choice between case-by-case approaches that set legal norms for each firm individually as opposed to all firms collectively.

The economic analysis of regulatory procedures and enforcement would improve by taking account of trade-offs that have been made between, for example, participatory rights, flexibility in the pursuit of regulatory objectives and accountability, or between procedures of due process and ease and effectiveness of enforcement. The analysis may then be taken even one step further, and legal principles themselves may become the object of economic analysis: norms of accountability, equality and consistency of treatment may be analysed as instruments for promoting commitment, limiting rent-seeking and preventing corruption.\(^{18}\)

Besides improving the explanatory and predictive power of economic theories, combined legal and economic research would enhance the quality of economic evaluations and the quality and effectiveness of expert economic advice on existing or proposed regulatory procedures, institutions and instruments. Decision-making, implementation and transformation of those regulatory regimes almost always depends on an assessment of legal, economic and political criteria.

---

\(^{18}\) Levy and Spiller (1994); Rose-Ackerman (1999).
6.4 Future research

Some important changes have recently taken place in the field of economic regulation, particularly with respect to the network industries: gas, electricity, telecommunications, water and sewerage, railways and suchlike. Traditionally in Europe, these industries have been organized as public enterprises, but they have been transformed into private enterprises and the industries have been opened up for competition. As public enterprises, these firms integrated the production, transport and delivery of services in one firm. The process of liberalization and the introduction of competition has often caused the network to become separated from the other parts of the industry. Public enterprises as governance structures of network industries have been substituted by the economic regulation of the networks and of the privatised former monopolists. This development raises a number of interesting empirical and theoretical research questions, from both a normative and a positive perspective.

First of all, if a comparison is made between the public enterprises as they were and the privatised industries as they are now, can it be said that the overall performance of the industries has been improved by this change in governance, or differs the change in performance of the different networks industries? Which specific aspects of performance have changed: have prices been driven down or has the effect been limited to costs, and what changes have taken place in the degree of organizational and technological innovation? What are the driving forces of the differences in strategies of transformation and the introduction of competition in the different sectors: why is it that sometimes the network remains under the control of the former monopolists while at other times the network is separated from the rest of the industry and brought under public control? What are the consequences for the performance of the industries of separating versus integrating the networks with the production or the delivery of services?

Secondly, unlike in the US where rate of return regulation has been more common, the dominant form of regulation in Europe has been by price capping. Has price capping led to lower prices and costs compared with rate of return regulation? Fair rate of return regulation makes the returns from investments more certain and credible compared with price cap regulation. This is because, although price cap regulation, may offer a higher potential profitability, profits
are less secure and investors will fear opportunistic behaviour by regulators, who may be tempted to extract unexpected profits by increasing the x-factor in the pricing formulae. Is the hypothesis of opportunistic behaviour confirmed, and, if so, what are the implications for the efficiency of investments under this form of regulation? Furthermore, given that there are incentives for overinvestments with rate of return regulation, which of the two types of regulation performs best with respect to dynamic efficiency? Under a given price cap, firms will be able to make extra profits by cutting the quality of services (safety, reliability, durability, frequency etc.). How can quality be made objectively measurable, what are the long-term developments in quality dimensions, and do these developments follow consumer’s preferences? Given the measurement difficulties that exist, what are the efficient ways of preventing the downgrading of difficult-to-measure quality aspects and how does this compare with the approaches taken by regulators?

Finally, the former public enterprises were required to pursue certain social policy objectives. For example, these enterprises had to meet all reasonable demand, including in rural areas, they had to provide public call boxes, a number of post offices per region, a prescribed number of stopping places at prescribed intervals, a given frequency of arrivals and departures per hour, and a minimum number of railway stations. At the same time, these firms were prohibited from using contract terms (prices) to discriminate unduly against certain groups, including customers in rural areas. Furthermore, separate provisions and facilities were supplied for special consumer groups such as deaf, sick, disabled or blind people. These social objectives were financed by a combination of monopoly pricing, cross-subsidies and subsidies from the treasurer. However, monopoly pricing and cross-subsidies are incompatible with competition, and subsidies are not allowed within European regulations. The question arises as to how these social policy objectives can best be achieved and what has been done in practice to pursue them? Is it efficient and effective to place a universal service obligation on the former monopolist and to finance these obligations partly through access deficit contributions from other operators using the network, or through the creation of a universal service fund19? If social provisions

---

19 As has been suggested by the 1997 EU Directive on Interconnection and Universal Service (97/33 EC).
for special consumer groups are not part of the universal service requirement, what other provisions for these groups have been made to guarantee adequate access to the services?

Another line of future research pertains to the field of social regulation. Although not as conspicuous as the developments with the regulation of network industries, significant changes are also taking place in the Netherlands with respect to health, safety and environmental regulation. The justification for these changes has been the perception that complying with social regulation has disproportionately increased the cost for the industries. Specifically, there are changes in the choices of instruments to attain the goals of social regulation. For example, in the field of environmental regulation, the current policy objective is that at most 25%, down from 45%, of businesses should require a licence before engaging in an activity, with general regulatory standards sufficing for the remaining businesses. In the field of health and safety regulation, rather than using specific regulatory standards that incorporate quantifiable units, the current principle regulatory policy is to pursue these targets through general regulatory target standards and legally non-binding covenants as concluded by the organizations of employers and employees at industry level.

The choice of instruments in the field of social regulation or the changes therein have not been subjected to a regulatory law & economics analysis. More generally, unlike much economic and social regulation in the United States, legislation and regulations in the Netherlands are not subjected to an economic, or, more specifically, a law & economics, appraisal. This situation raises a number of questions. Is it true, as the changes in policy suggest, that the costs of regulation outweigh the benefits, or that even if the benefits exceed their costs, even greater benefits can be realized by substituting other forms of regulation? Is it true, as has been suggested of social regulation in the United States, that a reallocation of resources could double the health gains in terms of lives that could be saved? What can be said about the costs per life saved and other health effects of the existing social regulations in the Netherlands and in Europe, and is it possible to prevent more accidents and save more lives at no additional resource costs? If the current social regulations are characterized by inconsistencies and inefficiencies, what are the causes of these deficiencies and what would the ideal risk-regulating institutions look like? If politicians and regulators
overreact to public demand in the wake of accidents and catastrophes by taking direct and visible, but not necessarily effective, measures, could a regulatory cost-benefit analysis contribute to a more systematic and rational form of policymaking? What are the strengths and weaknesses of regulatory appraisals founded on law & economics analyses, and how should uncertainties, equity considerations and distributional consequences be dealt with? What are the specific institutional requirements for safeguarding regulatory appraisals from inappropriate interest group pressure, and should independent expert agencies perform the regulatory appraisals of alternative instruments? Given that governments respond to the demands of the public for regulation and given the public’s perception of risks, how should one inform the interest groups concerned, and organize public involvement? Extensive regulatory appraisals will be a costly form of decision analysis: is there a way of deciding which regulations should be subjected to such an analysis, should one differentiate between different kinds of regulatory initiatives and reserve only the most significant of them for regulatory appraisals, or would other institutions be better?
REFERENCES


Ashenfelter, O. and R. Layard (eds.) (1987), Handbook of Labor Economics, Amsterdam, North Holland.


Bakker, Wieger en Frans van Waarden (red.) (1999), Ruimte rond regels. Stijlen van regulering en beleidsuitvoering vergeleken, Boom.


Barzel, Yoram (1985), 'Transactions Costs: Are They Just Costs?', Journal of Institutional and Theoretical Economics 141 (no.1), 4-16.


References


PUBLIC AND PRIVATE INTERESTS IN REGULATION

Bohm, Peter (1987), Social efficiency, London, Macmillan Education LTD.


Buchanan, James M., Tollison, Robert D, Tullock, Gordon (1980), Toward a Theory of the Rent-Seeking Society, College Station, Texas A. & M. University Press


References


Cheung, Steven N.S. (1980), The Myth of Social Cost, Cato Institute, San Fransisco, California.


PUBLIC AND PRIVATE INTERESTS IN REGULATION


Gruenspecht, Howard K. and Lave, Lester B (1989), 'The economics of health, safety and environmental regulation', in: Schmalensee, Richard and Willig,

168
References

Robert D., Handbook of industrial organization II, Amsterdam, North Holland, 1508-1550.


KNMG (1984), Gedragsregels voor artsen, KNMG, Utrecht.

KNMG (1981), Statuten en huishoudelijk reglement, Utrecht.


Landelijke Specialisten Vereniging der KNMG (1979), Model-associatiecontract voor medische specialisten, LSV, Utrecht.

Landelijke Specialisten Vereniging der KNMG (1990), De specialist van morgen, LSV, Utrecht.


Landelijke Specialisten Vereniging der KNMG (1990), 'Richtlijnen bepaling hoogte goodwillsom', Medisch Contact, nr. 41.
PUBLIC AND PRIVATE INTERESTS IN REGULATION


References


Mijn, W.B. van der (1989), *Beroepswetgeving in de gezondheidszorg*, Kluwer, De- 

venter.


Ng, Yew-Kwang (1985), 'Some Fundamental Issues in Social Welfare', in: Fei- 


PUBLIC AND PRIVATE INTERESTS IN REGULATION


Olson, Mancur (1982), The rise and decline of nations, New Haven, Yale University Press.

References


Pensioenkamer (1989), Pensioenkaart van Nederland, Den Haag.


Pitlo, A. (1979), Algemeen deel van het verbintenissenrecht, Gouda Quint, Arnhem.

PUBLIC AND PRIVATE INTERESTS IN REGULATION


Post, Doede (ed.) (1990), Hier meer, daar minder, Meppel, Ziekenfonds ‘Het Groene Land’.


Prijzenwet, 1961, Staatsblad 135.


References


PUBLIC AND PRIVATE INTERESTS IN REGULATION


PUBLIC AND PRIVATE INTERESTS IN REGULATION


References


Wet betreffende de verplichte deelneming in een bedrijfspensioenfonds, Wet van 17 maart 1949, Staatsblad J 121, art. 3.

Wet economische mededinging, 1958, Staatsblad 413.


Wet regelende de uitoefening van de geneeskunst, 1865, Staatsblad 60.

Wet tarieven gezondheidszorg, 1980, Staatsblad 646.

Wet van 25 december 1878, Staatsblad 222, houdende regeling der voorwaarden tot verkrijgen der bevoegdheid van arts enz.

Wet voorzieningen gezondheidszorg, 1982, Staatsblad 563.


Wittman, Donald (1977), Prior regulation versus post liability: the choice between input and output monitoring, Journal of Legal Studies 6, 193-211.


Wolf, Charles (1993), Markets or Governments, choosing between imperfect alternatives, Cambridge Massachusetts, MIT Press.

Ziekenfondswet, 1964, Staatsblad 392.

Summary in Dutch

PUBLIEKE EN PRIVATE BELANGEN IN REGULERING

Rechtseconomische reguleringsopstellen

Inleiding


Er kunnen drie typen van rechtseconomische analyses van reguleringen worden onderscheiden, die echter in de praktijk niet altijd goed te scheiden zijn. Allereerst kan de rechtseconomische analyse gericht zijn op de verklaring van regulering en de toepassing van specifieke instrumenten op bepaalde terreinen en sectoren. Waarom is de toetreding tot de markt voor medische verrichtingen niet vrij en waarom vindt prijs- en kwaliteitsregulering plaats? Kan de regulering verklaard worden uit de behartiging van het
algemeen belang of zijn regels nadelig voor de welvaart en vinden zij hun oorsprong in het nastreven van deelbelangen? In de tweede plaats kan de analyse van rechtsregels gericht zijn op de bepaling van de effecten: zijn de regels effectief, zijn er onverwachte of zelfs averechtse effecten? In de derde plaats kan de rechtseconomische analyse gericht zijn op een evaluatie van de rechtsregels. Hoofdstukken twee tot en met vijf bevatten hoofdzakelijk verklarende analyses van rechtsregels of ontwikkelingen in reguleren. Hoofdstukken twee en vijf bevattend daarnaast op onderdelen ook een effectanalyse en een evaluatie. In hoofdstuk één wordt een overzicht en synthese gegeven van de economische theorieën op het gebied van reguleren. Het laatste hoofdstuk geeft een samenvatting en conclusies. De rechtseconomische analyses in dit boek maken gebruik van ‘public and private interest’ theorieën van reguleren.

**Hoofdstuk 1 Reguleringstheorieën**

In dit hoofdstuk wordt een overzicht, kritiek en synthese gegeven van de op verklaring gerichte reguleringstheorieën. In de literatuur worden twee grote groepen verklaringstheorieën van reguleren onderscheiden, de publieke en private belangentheorieën. Volgens de publieke belangentheorieën kunnen wettelijke en bestuursrechtelijke regels verklaard worden uit het nastreven van het publieke belang. Dit publieke belang bestaat uit de bevordering van een efficiënte allocatie van middelen die aangewend worden voor de productie van individuele en collectieve goederen en diensten. Onder bepaalde omstandigheden slagen markten er niet in om deze efficiënte allocatie tot stand te brengen. Markten kunnen instabiel zijn, of ontbreken, gekenmerkt worden door onvolledige mededeling of onwenselijke uitkomsten genereren. Indien privaatrechtelijke alternatieven inefficiënt zijn kunnen wettelijke of bestuursrechtelijke regels bijdragen aan een efficiëntere allocatie voorzover de reguleringskosten geringer zijn dan de toename van de welvaart die voortvloeit uit de toepassing van de regels. In de literatuur is op verschillende aspecten van deze publieke belangentheorieën kritiek gekomen. Omdat er meerdere omschrijvingen bestaan van ‘publieke belangen’ zijn theorieën moeilijk weerlegbaar. Vervolgens ontbreken in de theorieën verbanden die aangeven hoe marktonvolkomenheden corrigerende wetgevende en bestuurlijke activiteiten genereren. Ook lijken veel waarnemingen in strijd met
Ook lijken veel waarnemingen in strijd met voorspellingen die uit de theorieën voortvloeien. De publieke belangentheorieën zouden eigenlijk normatieve, evaluerende theorieën zijn, die gebruikt worden als positieve, verklaringstheorieën van regulering. De publieke belangen theorieën geven verder weinig inzicht in de verdeling van de voor- en nadelen van de reguleringen over relevante groepen. Dat is merkwaardig omdat verdelingsvraagstukken vaak centraal staan bij de bepaling van het publieke belang.


In de literatuur is een aanzet gedaan tot het beter met elkaar in overeenstemming brengen van de publieke en private belangentheorieën van regulerings. Deze aanzet is nog in ontwikkeling.
Hoofdstuk 2  De verklaring van deregulering in Nederland

In dit hoofdstuk worden de Chicago-reguleringstheorieën, de publieke belangen reguleringstheorieën en politiek-economische reguleringstheorieën toegepast om de effectuering van het dereguleringsbeleid in Nederland te verklaren. De Chicago-reguleringstheorieën vormen het fundament en belangrijkste bestanddeel van de private belangen reguleringstheorieën. Sinds het begin van de jaren tachtig heeft deregulering van economische activiteiten op de politieke agenda gestaan maar tot een werkelijke vermindering van de aard en omvang van regulering heeft dat niet geleid. Vanaf de tweede helft van de jaren negentig komt er meer vaart in de bevordering van marktwerking en de vermindering van regulering van economische activiteiten. Welke verklaringen kunnen hiervoor gevonden worden?

Uit de publieke belangen reguleringstheorie kunnen verschillende hypothesen worden afgeleid over deregulering van voorheen gereguleerde activiteiten. Vraag- en/of aanbodfactoren kunnen de marktimperfecties die aanleiding gaven tot regulering, hebben opgeheven. Ook is het mogelijk dat de relatieve informatie-, transactie- en organisatievoordelen van overheidsregulering verminderd zijn ten opzichte van die van de private sector.

Vanuit het gezichtspunt van de Chicago-reguleringstheorieën kunnen tenminste vier mogelijke oorzaken worden genoemd voor de vermindering van overheidsregulering van economische activiteiten. In de eerste plaats kunnen belangengroepen de overtuiging gekregen hebben dat deregulering of zelfregulering hun doelen beter dient dan regulering. In de tweede plaats is het mogelijk dat de politieke effectiviteit van benadeelde groepen is vergroot en/of dat de informatie- en organisatieadviezen van benadeelde groepen van regulering verminderd zijn. Ook is het mogelijk dat allerlei factoren ertoe hebben geleid dat wijzigingen zijn aangebracht in de relatieve politieke ondersteuning die belangengroepen kunnen bieden. Tenslotte, wanneer de welvaartsnadelen van regulering in de loop der tijd accumuleren, kan de balans ten gunste van de status quo worden ondermijnd.

Op basis van politiek-economische reguleringstheorieën kunnen de volgende oorzaken van deregulering gevonden. In de eerste plaats kan een verandering in de balans van belangengroepen resulteren in deregulering. Vervolgens is het mogelijk dat politieke ondernemers barrières slechten die de bestaande struc-
tuur instandhouden. In de derde plaats kunnen politici succesvol politieke ondersteuning zoeken voor dereguleringsvoorstellen door het electoraat van informatie te voorzien over bijvoorbeeld de welvaartsnadelen en welvaartsverdrachten. Ook is het mogelijk dat regelgevers juist de complexiteit van reguleringsvraagstukken gebruiken ten einde het electoraat op ideologisch wijze te overtuigen dat deregulering de welvaart vergroot.

Uit een toepassing van bovengenoemde oorzaken op de organisatie, omstandigheden en resultaten van deregulering in Nederland kan geconcludeerd worden dat in het bijzonder twee variabelen hebben bijgedragen aan de effectiviteit van het dereguleringsbeleid: publieke opinie en politiek ondernemerschap. De eerste variabele zorgde voor een verandering in de machtsbalans van de politieke partijen: na de verkiezingen van 1994 maakten de christendemocratische partijen voor het eerst sinds 1918 geen deel meer uit van de regeringscoalitie. De tweede variabele, het politieke ondernemerschap, bewerkstelligde het slechten van bestaande belemmeringen in de politiek-ambtelijke besluitvormingsstructuur. De voormalige voorzitter van de bijzondere commissie Deregulering van overheidsregelingen werd benoemd tot secretaris-generaal van het departement van Economische Zaken en kon in die hoedanigheid effectief de bureaucratie besturen en het in zijn optiek noodzakelijke dereguleringsbeleid bevorderen.

**Hoofdstuk 3 Regulering en zelfregulering van de medisch specialisten in Nederland**

In dit hoofdstuk wordt onderzocht of de regulering en zelfregulering van medisch specialisten verklaard kan worden uit het nastreven van het algemeen belang of uit het nastreven van particuliere belangen. De markt voor verrichtingen van medisch specialisten is sterk gereguleerd ten aanzien van toetreding, marktverdeling, kwaliteit en tarieven. De toetreding tot de beroepsgroep is beperkt door de numerus fixus en de lange duur van de opleiding: na het afleggen van het artsexamen moet men zich specialiseren tot huisarts, sociaal-geneeskundige of medisch specialist. Deze specialisaties zijn een privaatrechtelijke regeling van de KNMG. Door opname in het registratieregister beperkt de arts de algemene medische bevoegdheid en wordt een zekere monopoliepositie
verkregen op het gebied van het specialisme. De titel- en beroepsbescherming schermen de markt volledig af. Andere toetredingsbeperkingen zijn afkomstig van de overheid, de verzekeraars en van de beroepscode en de gedragsregels van de artsenorganisaties zelf. Eenmaal toegelaten tot de markt, wordt de onderlinge rivaliteit beperkt door regels terzake van prijzen, kwaliteiten en hoeveelheden. De tarieven komen tot stand door overleg tussen de beroepsverenigingen van specialisten en de verzekeraars. Deze tarieven moeten worden goedgekeurd door een publiek orgaan (COTG). Deze tarieven zijn in principe maximumtarieven, maar in de beroepscode van de artsen is bepaald dat de arts zich dient te houden aan de landelijk door de beroeporganisaties overeengekomen tarieven. Ook vormen van marktverdeling worden toegepast: zo is bepaald dat artsen terughoudend moeten zijn bij het overnemen van patiënten in eenzelfde werkgebied en is het verboden binnen zes maanden eigen patiënten in de praktijk op te nemen na hulpverlening tijdens waarneming of controle. Kwaliteitsbewaking vindt vooral plaats door middel van structuren (opleidingssteun, organisatie van de medische staf) en weinig door de toetsing van de uitkomsten van het medisch handelen zelf. Rivaliteit is vooral afkomstig van de georganiseerde alternatieve geneeskunde (homeopathie, acupunctuur).

Kunnen de regels verklaard worden uit het nastreven van het algemeen belang, dat wil zeggen, liggen marktimperfecties en doelmatigheidsoverwegingen ten grondslag aan de regulering en zelfregulering? Medische verrichtingen zijn ervaringsgoederen en vertrouwensgoederen. Zonder regulering zou vrije toetreding van kwakzalvers mogelijk worden en averechte selectie optreden. Een opleidingsstelsel uitgevoerd door de beroepsgroep zelf en vaste prijzen garanderen een minimum kwaliteitsniveau. Zelfregulering is efficiënt omdat artsen terzake deskundiger zijn, omdat veranderingen gemakkelijker zijn te effectue ren en omdat regelgeving en handhaving door de beroepsgroep zelf doelmatiger is. Aan de vraagkant is het verzekeringstelsel opgezet teneinde toegankelijkheid te waarborgen. Vanwege de risico’s van ‘moral hazard’ is het vervolgens wel noodzakelijk grenzen te stellen aan vrije toetreding. Omdat de tarievenstructuur en de marktverdeling vastliggen, is regelgeving ontwikkeld om het inkomensniveau en de inkomensverschillen te beperken.

Niettemin zijn er waarnemingen die strijdig lijken te zijn met de algemene belangenthese van regulering. Incidentele registraties van de uitkomsten van het medisch handelen laten zien dat er onverklaarbaar grote verschillen bestaan
in verrichtingen per medisch specialist, ook wanneer rekening gehouden met verschillen in bevolkingssamenstelling. Het administratieve tarievenstelsel leidt er verder toe dat de verhouding tussen tarieven en reële kosten verstoord is. Ook is het inkomensniveau van medisch specialisten hoog, zij behoren tot de 1% best verdienende zelfstandigen en ook bestaan er grote verschillen tussen inkomensniveau tussen de verschillende specialismen. Tenslotte lijken regulering en zelfregulering verder te gaan dan noodzakelijk is op grond van de geconstateerde vormen van marktimperfecties.

Deze waarnemingen passen ook beter bij de private belangentheorie van regulering. Volgens deze theorie is regulering en zelfregulering te verklaren uit het tot stand brengen van inkomens- en welvaartsoverdrachten in ruil voor politieke ondersteuning. Met name kleine groepen, strak georganiseerd rondom een beperkt beleidsveld, hebben de meeste kans van slagen. Succes wordt verder bevorderd indien de benadeelde groep omvangrijk en afhankelijk is en indien informatiekorten aanwezig zijn. De markt voor medische verrichtingen en de groep van medisch specialisten voldoen aan deze kenmerken. Medisch specialisten betreft een kleine groep betrokkenen (7500 personen) die goed georganiseerd zijn in de verschillende wetenschappelijke verenigingen. Tussen patiënten en artsen is de informatie over de verrichtingen en over de kosten bovendien sterk asymmetrisch verdeeld. Eventuele premiestijgingen die resulteren uit een stijging van de tarieven, worden verspreid over miljoenen premiebetalers. De ontwikkeling van tegenmacht is onder deze omstandigheden niet te verwachten.

Verschillende vormen van marktfalen en marktimperfecties lijken een verklaring te kunnen bieden voor de gevonden vormen van regulering en zelfregulering. Niettemin zijn er waarnemingen die met deze verklaring in strijd zijn. Deze waarnemingen komen weer goed overeen met de rivaliserende verklaring voor regulering en zelfregulering die afkomstig is uit de particuliere belangengroepentheorie van regulering.
Hoofdstuk 4 Omgekeerde solidariteit in pensioenstelsels

Veel Nederlandse pensioenstelsels worden gekenmerkt door omgekeerde solidariteit, dat wil hier zeggen dat degenen die geen carrière maken bijdragen aan de financiering van hen die wel carrière maken. In de praktijk betekent dit dat overdrachten vooral plaatsvinden van vrouwen naar mannen en van lager betaald naar hoger betaald personeel. In dit hoofdstuk wordt onderzocht hoe omgekeerde solidariteit ontstaat en welke verklaring ervoor te vinden is. Na een onderzoek van verschillende soorten pensioenstelsels blijken de ‘verborgen’ overdrachten plaats te vinden in eindloonstelsels, waarbij iedereen eenzelfde percentage van het laatst verdiende loon als pensioen ontvangt, bijvoorbeeld 70% en ieder eenzelfde percentage van het inkomen aan premie betaalt, bijvoorbeeld 1,75% per jaar.

Er zijn verschillende redenen aan te geven waarom betrokkenen een relatief kostbaar pensioencontract afsluiten. Pensioencontracten zijn standaardcontracten waarvan werknemers bij het aangaan van de arbeidsovereenkomst meestal niet op de hoogte zijn, ze zijn ingewikkeld door het gespecialiseerde (juridische) taalgebruik en abstract omdat zij voor vele uiteenlopende gevallen bedoeld zijn. Alternatieve pensioencontracten zijn bovendien lastig disconteerbaar, vergelijkbaar en evaluateerbaar en tenslotte ligt de pensioengerechtigde leeftijd vaak ver weg. Maar bovendien is feitelijk een situatie ontstaan waarin de individuele werkgever en de individuele werknemer weinig invloed meer kunnen uitoefenen op de inhoud van de pensioenovereenkomst. Wettelijk is bepaald dat de werkgever de uitvoering van de pensioentezegging moet overdragen aan (bijvoorbeeld) bedrijfspensioenfondsen. Voor de werkgever is de werkgeversorganisatie de onderhandelingspartner van het bedrijfspensioenfonds. Bovendien wordt op verzoek in meer dan 80% van de gevallen deelname aan het pensioenfonds door de minister verplicht gesteld, ook voor niet bij een werkgeversorganisatie aangesloten werkgevers. Voor werknemers is de vakbeweging de onderhandelingspartner op het gebied van de pensioenen en heeft de ondernemingsraad instemmingsrecht bij het instellen en wijzigen van pensioenverzekeringen. Wettelijk is tenslotte bepaald dat afkoop van de pensioenrechten bij het pensioenfonds verboden is.

De reden dat overdrachten van niet-carrièreakers naar carrièreakers plaatsvinden hangt samen met de wederzijdse belangen van het pensioenfonds en de
werkgevers- en werknemersorganisaties. Voor het pensioenfonds vereenvou-
digt het de administratie wanneer een doorsnee premie geheven wordt in plaats
van actuarieel juiste premies. Voor werkgevers betekent een eindloonstelsel met
uniforme premies zowel een stimulans voor werknemers om zich in te zetten
als een methode om personeel te screenen op karakteristieken als mobiliteit en
geschiktheid voor hogere functies. Doordat bovendien de verschillen tussen
hoog en laag betaald personeel optisch verkleind worden, versluipt het pensi-
oensysteem de beloningsstructuur. Dat vermindert de kans op arbeidsonrust.
Werknemersorganisaties zullen zulke pensioensystemen niet bestrijden. Als
democratische organisatie zal de vakbond de voorkeuren van de mediane
werknemer weerspiegelen. Deze werknemer zal ouder zijn en dichter bij de
pensioengerechtigde leeftijd en zodoende een voorkeur vertonen voor een eind-
loonstelsel. Eenzelfde voorkeur zal resulteren indien vakbondsbestuursleden
meer ruimte hebben de eigen voorkeuren te effectueren: hun leeftijd zal niet
veel afwijken van die van het mediane lid en eerder nog hoger liggen. Verder
zal een collectief pensioenstelsel de macht van de vakbond aanzienlijk doen
toenemen. Wettelijk is bepaald dat zij de helft van de bestuurszetels van het
pensioenfonds mogen innemen. Op deze wijze krijgen zij ook invloed op de in-
vesteringen gefinancierd uit bijdragen van niet-vakbondsleden. Niet alleen
geeft dit meer prestige en invloed aan vakbonden, het is ook te verwachten dat
naar medezeggenschap strevende werknemers vaker lid zullen worden van de
vakbond. Met uniforme premies wordt tenslotte de concurrentie uitgeschakeld
en wordt een mogelijke neerwaartse druk op de loonstructuur voorkomen.
Een internationaal vergelijkend onderzoek naar de pensioensystemen van
Duitsland, Frankrijk, België en Engeland toont aan dat sommige van de pensi-
oensystemen van België en Engeland eveneens gekenmerkt worden door om-
gekeerde solidarity.

**Hoofdstuk 5 Het concurrentiebeding: onredelijk of efficiënt?**

In Nederland sluit één op de vijf werknemers een arbeidsovereenkomst waarin
een concurrentiebeding is opgenomen. Dat is een beding waarbij de werkgever
en de werknemer overeen komen dat de werknemer aan het einde van de ar-
beidsovereenkomst gedurende een bepaalde periode niet op een bepaalde wijze
werkzaam zal zijn, bijvoorbeeld niet bij een soortgelijk bedrijf, niet in dezelfde regio enz. Regelgevers en rechters zien deze overeengekomen bedingen veelal als een uitdrukking van de ongelijke machtspositie van de werknemer. Soms worden zulke bedingen eenvoudig verboden (California, Mexico), soms zijn er wettelijke beperkingen (België, Duitsland) en soms heeft de rechter al of niet vergaande matigingsmogelijkheden (bijvoorbeeld Nederland). In Nederland is een wetsvoorstel ingediend waarbij de toepassing van het beding aan strakke voorwaarden gebonden wordt.

In dit hoofdstuk wordt onderzocht of een concurrentiebeding inderdaad uitdrukking is van een ongelijke machtspositie of dat een dergelijk beding efficiënt is en bijdraagt aan de welvaart van de contractpartijen. Ook onder economen is het lange tijd gangbaar geweest om bij ongewone contractsvoorwaarden te zoeken naar een monopolieverklaring. Pas vanaf het begin van de jaren zeven- tig zijn er stromingen ontstaan (Rechtseconomie, Institutionele economie) die zoeken naar alternatieve (efficiëntie)verklaringen voor ongewone, veel voor- komende contractsbepalingen.

Het is twijfelachtig of de juridische theorieën over de oorzaken van het concurrentiebeding stand kunnen houden. Machtsposities komen minder voor, het concurrentiebeding komt ook voor in concurrerende sectoren en tenslotte is niet waarschijnlijk dat informatietekorten ten grondslag liggen aan de verbreiding van het beding. Het concurrentiebeding komt vooral voor bij die functies waarvan men mag verwachten dat de betreffende werknemers voldoende geschoold en/of ervaren zijn om op de hoogte te zijn van de contractsvoorwaarden. Vier theorieën worden besproken die de toepassing van het concurrentiebeding zou kunnen verklaren. Werkgevers zijn vaak niet goed op de hoogte van relevante kenmerken van werknemers zoals bijvoorbeeld de mate van investeringsbe- reidheid of flexibiliteit. Het concurrentiebeding kan door werkgevers worden toegepast om een betrouwbare signaal te krijgen van solliciterende werknemers over hun beoogde duurzame betrokkenheid bij de onderneming. Een dergelijk beding is efficiënter dan een lange termijn arbeidscontract. Het concurrentiebe- ding wordt hier dus toegepast als instrument tot zelfselectie van werknemers.

In de tweede plaats kan een concurrentiebeding door werkgevers worden ge- bruikt om verrichte opleidingsinvesteringen veilig te stellen. De analyse laat zien dat het voor werkgevers doelmatiger is voor een dergelijk doel een beding met betrekking tot de terugbetaling van opleidingskosten contractueel overeen
Summary in Dutch

te komen. In de derde plaats kan een concurrentiebeding overeen gekomen worden om bedrijfsgeheimen en vertrouwelijke informatie te beschermen. Vanwege bewijsrechtelijke problemen kan onder bepaalde omstandigheden een concurrentiebeding efficiënter zijn dan een geheimhoudingsclausule. In de vierde plaats kan het concurrentiebeding worden aangewend om risicovolle en kostbare heronderhandelingen over de verdeling van de opbrengst uit de arbeidsrelatie te voorkomen. Zowel de werknemer als de werkgever kunnen in kwetsbare omstandigheden komen te verkeren wanneer zij ondernemingspecifieke investeringen hebben verricht. De partij die de investeringen verricht heeft staan bloot aan het risico van ‘hold-up’ door de andere partij; deze zal in de verleiding kunnen komen zich meer van de toegevoegde waarde van de arbeidsrelatie toe te eigenen dan aanvankelijk was overeengekomen. Omdat partijen zich bewust zijn van dit risico, zullen investeringen alleen plaatsvinden indien garantes tegen dit risico van ‘hold-up’ verkregen zijn. Het toepassen van het concurrentiebeding heeft nu tot gevolg dat een werknemer niet effectief kan dreigen met vertrek omdat de alternatieve werkgelegenheidsmogelijkheden beperkt zijn. Het concurrentiebeding beschermt het ontstaan en de overeengekomen verdeling van het surplus uit de arbeidsrelatie; dat is voordelig voor beide partijen: investeringen, economische groei en de werkgelegenheid en het loon niveau nemen toe.

Hoofdstuk 6 Samenvatting en conclusies

Hiervoor is een samenvatting gepresenteerd van de afzonderlijke hoofdstukken. In dit gedeelte worden overkoepelende conclusies getrokken en algemene beleidssimplicaties geformuleerd. In de literatuur is een debat gaande over de waardering van de publieke en private belangen reguleringstheorieën. Vooral de publieke belangen reguleringstheorieën worden sterk bekritiseerd. Om theorieën te kunnen beoordelen en kritiek te kunnen uitoefenen zijn beoordelingscriteria benodigd. Deze criteria zijn afkomstig uit de economische methodologie. Met behulp van zulke criteria zijn de belangrijkste kritiekpunten samengevat. Een eerste overkoepelende conclusie is dat het op basis van de bestaande stand van zaken in de ontwikkeling van de reguleringstheorieën, het niet mogelijk is één van beide groepen van
theorieën te rangschikken als de beste in termen van verklaringskracht. De ‘public and private interest’ theorieën van regulering kunnen nog het beste als grondtheorieën worden getypeerd waaruit specifieke modellen worden afgeleid voor de verklaring van reguleringscomplexen. Deze grondtheorieën worden aanvaard omdat zij geacht worden een aannemelijk beeld van de werkelijkheid te schetsen (Klant). Een tweede conclusie is dat de kracht van de ‘public and private interest’ theorieën dan ook vooral ligt in het gebruik van deze theorieën als parallelle en tegenover elkaar staande denkkaders en analyseschema’s. Juist door het gebruik van bêide analysekaders kan duidelijker worden onder welke omstandigheden een bepaalde reguleringstheorie eerder van toepassing is dan een alternatieve en kunnen bijvoorbeeld nieuwe hypothesen ontwikkeld worden over de bijdrage van zulke reguleringen aan de ontwikkeling van de economische groei en de welvaart. De waarde van de ‘public and private interest’ theorieën van regulering ligt naast de bijdrage aan de theorieontwikkelingen ook in de toepassingsmogelijkheden voor de praktijk. Niet alleen kunnen bepaalde waarnemingen beter begrepen worden door toepassing van beide denkkaders, ook kunnen de theorieën behulpzaam zijn bij voorstellen voor toepassing van reguleringen in de praktijk. Meer in het algemeen kan de toepassing van de publieke belangentheorieën bijdragen aan het ontdekken van mogelijke vormen van marktfalen en efficiënte reguleringsvormen, en kan het toepassen van de inzichten van de private belangentheorieën mogelijke reguleringsvoorstellen minder kwetsbaar maken voor misbruik door belangengroe- pen.

De beleidsimplicaties liggen op de specifieke terreinen van onderzoek en op de overkoepelende conclusies. Hier zullen een aantal overkoepelende beleidsimplicaties verwoord worden. De discussie over reguleringstheorieën heeft laten zien onder welke voorwaarden en omstandigheden inefficiënte pressie van belangengroepen meer plausibel is. Een eerste implicatie is dat het nastreven en behartigen van publieke belangen met zich meebrengt dat relevante belangen geïdentificeerd en publiek gemaakt worden en dat breed informatie verspreid wordt over de verdeling van kosten en baten van reguleringsvoorstellen. Een tweede implicatie van de wederzijdse confrontatie van reguleringstheorieën betreft het niveau en de aard van de besluitvorming. Bij de bepaling van het optimale niveau van de besluitvorming moet rekening gehouden worden met
rent-seeking door belangengroepen. Ten gevolge van de kosten van informatie en organisatie verschillen deze risico’s voor de verschillende niveaus van besluitvorming. Onder bepaalde voorwaarden zal dan bijvoorbeeld eerder de voorkeur gegeven kunnen worden aan algemene regels boven vergunningsvoorschriften met geïndividualiseerde voorwaarden.

Een volgende implicatie betreft de institutionele context van het reguleringsproces. De voorbereiding van wet- en regelgeving is tot op heden vooral het domein van (wetgevings- en bestuurs)juristen. Het verdient aanbeveling de samenwerking tussen juristen en specialisten op het terrein van de rechtseconomie te intensiveren. Reguleringsvoorstellen zullen op betere analyses van de werking van markten en regels gestoeld kunnen worden, waardoor de risico’s van averechtse effecten en onverwachte nevengevolgen van regelgeving verminderen. Ook op wetenschappelijk terrein zou de samenwerking tussen juristen en (rechts)economen versterkt moeten worden. De verklaringskracht van de economische theorieën zal kunnen toenemen en de kwaliteit en de effectiviteit van de economische advisering zal vergroot worden: besluitvorming over regelgeving is vrijwel altijd afhankelijk van juridische, economische en politieke overwegingen.