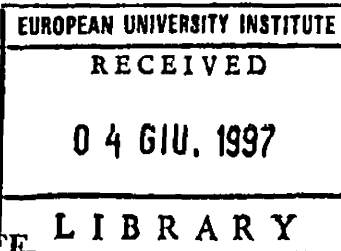EUROPEAN UNIVERSITY INSTITUTE
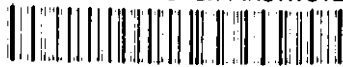**Department of Economics**

Research and Development,
Product Differentiation and Robust Estimation

Jeroen Hinloopen

Thesis submitted for assessment with a view to obtaining
the Degree of Doctor of the European University Institute

Florence, May 1997

B/C →D

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Research and Development, Product Differentiation and Robust Estimation

Jeroen Hinloopen

The Thesis Committee consists of:

Prof. Paul Geroski, London Business School
" Stephen Martin, EUI and University of Copenhagen, Supervisor
" Louis Phlips, EUI
" David Ulph, University College London
" Spyros Vassilakis, EUI

To all who taught me

Contents

Chapter 2   Subsidizing cooperative and noncooperative R&D in a differentiated oligopoly

Chapter 3   On the equivalence of subsidized cooperative and noncooperative R&D

II   Product Differentiation

Chapter 4 Spatial competition with an outside good

III   Robust Estimation

Chapter 5 Generalized M estimation

Chapter 6 Empirical research in Industrial Organization: a re-interpretation

References

x

Preface

The reader will notice that the present study consists of three parts which appear to be quite distinct. After turning the last page the persevering reader might still fail to see the link between the three themes. In that case I hope that reading the thesis has been rewarding anyway. Yet, there will be others who anticipate the connection between the three movements for which the present dissertation lays the foundation. Those I would like to invite to await subsequent writings, or to pursue the suggested research themselves.

A Ph.D.-thesis is never written without assistance, and the present text is no exception. In that respect the first person I want to thank is my supervisor, Professor Stephen Martin. His clear guidance, his continuous strong support, and his critical assessment of my work have made the writing of this dissertation a very pleasant experience.

My intellectual debt to my friend and co-author, Rien Wagenvoort, is especially visible in Part III of this dissertation. Working with him has been very rewarding, both on a personal and professional level, and I am looking very much forward to our future cooperation.

The second part of the present work would not have been of the same level if I had not worked with Charles van Marrewijk. Our cooperative research has enabled me to write Chapter 4. The hospitality of the (then) department of International Economics of the Erasmus University Rotterdam in the summer and winter of 1994 is also gratefully acknowledged.

Professor Louis Phlips encouraged me very strongly to write my first working paper. This difficult exercise not only taught me how to write scientific articles, it also forced me to disentangle the difficulties associated with R&D-subsidies and concomitant taxes. The reader will find the result in Chapter 2.

The present text also benefitted from numerous comments and suggestions given by participants at various conferences, summer-schools and workshops. The comments of Professors Paul Geroski and Jacques-Francois Thisse were especially helpful.

Chapter 6 could not have been written without all those who generously provided me with their data. In that respect I want to thank Professors Harry Bloch, Paul Dunne, José Mata, Jaime de Melo, Willard Mueller, Nicholas Oulton, Everett Peterson, John van Reenen, Gavin Reid, Avi Weiss, Wesley Wilson, and Yishay Yafeh.

My stay at DG III of the European Commission contributed to my understanding of the gap between economic theory and political practise. But I still hope that members of Unit A5 will find in the present dissertation some thought provoking proposals.

Any time span of almost four years will constitute an important period in one's life. For me the last four years are no exception. My stay in Italy is not only the period in which I wrote my Ph.D.-dissertation, but foremost it is the period in which I met my girlfriend, Karin Westerbeek. What she means to me goes well beyond the importance of the following pages.

Jeroen Hinloopen
Le Caldine, Spring 1997

xi

Part I
Research and Development

I do not doubt that the eternal mother of new wisdom - time - which has revealed so much to us which was unknown to our ancestors, will also disclose to our descendants what we wanted to know, yet were unable to discover.

- John Wilkins -

1 Potential policies for private R&D

1.1 Introduction

Technological progress influences considerably the design of contemporary society. Compact disc players bear little resemblance with traditional pick-ups, mobile phones evoke vague memories of Bell's original device, typewriters induce nostalgic thoughts when installing the latest version of a text editor, and the Train Grand Vitesse has only the same family name as its great grandfather. The list of technological improvements realized in the twentieth century is well nigh inexhaustible.

When economists consider technological progress, they are inclined to *measure* its economic impact. One such measurable effect is the result of an important mechanism set in motion by successful research: innovations can increase factor productivity, which in turn could contribute to economic growth (the economic literature on this relationship is vast; for a recent survey see Korres [1996]). In Figure 1.1 this relation is illustrated. For nineteen industrialized countries, the percentage-change in total factor productivity is related to the corresponding percentage-change in economic growth, computed over three consecutive time periods (1960-1973, 1973-1979, and 1979-1988).[1] The positive relation between growth in total factor productivity and economic growth is evident. This is confirmed by the estimated relationship, drawn as the solid line.[2]

It is normally presumed that firms decide on the basis of a cost-benefit analysis whether to pursue technological advancements. In that respect motivations to invest in research and development (R&D) abound.[3] On a fundamental level, innovations can increase productivity (see e.g. Hall and Mairesse [1995]), which could be an adequate response to increased competition possibly from abroad (see e.g. Bertschek [1995]). Successful innovations can also lead to an increase in market share or to the birth of new markets, thence increasing profitability (see e.g. Geroski et al. [1993]). Further, building up a broad scientific base could act as a barrier to entry (see Geroski [1993]), thereby securing future profits to a certain extent. Finally, consistently investing in R&D enhances individual firms' ability to learn from competitors' research (see Cohen and Levinthal [1989]). In sum, innovative firms are more competitive.

---

[1] The countries considered are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, The Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, The United Kingdom and The United States. Figures for total factor productivity are taken from Crafts [1992], while those concerning GDP are provided by Eurostat.

[2] The estimated relationship reads: $dGDP = 0.82 + 0.48 \, dTFP$, where both estimated coefficients are highly significant. Incidentally, the GM estimator described in Chapter 5 is used for this econometric analysis. For more details, see European Commission [1996a].

[3] In this study, R&D refers to a wide range of activities. Schumpeter already distinguished three phases in the process of technological advancement: invention, innovation, and diffusion. R&D is assumed here to encompass the first two of these phases. For a detailed analysis of Schumpeter's division, see Stoneman [1983].

%-Change GDP
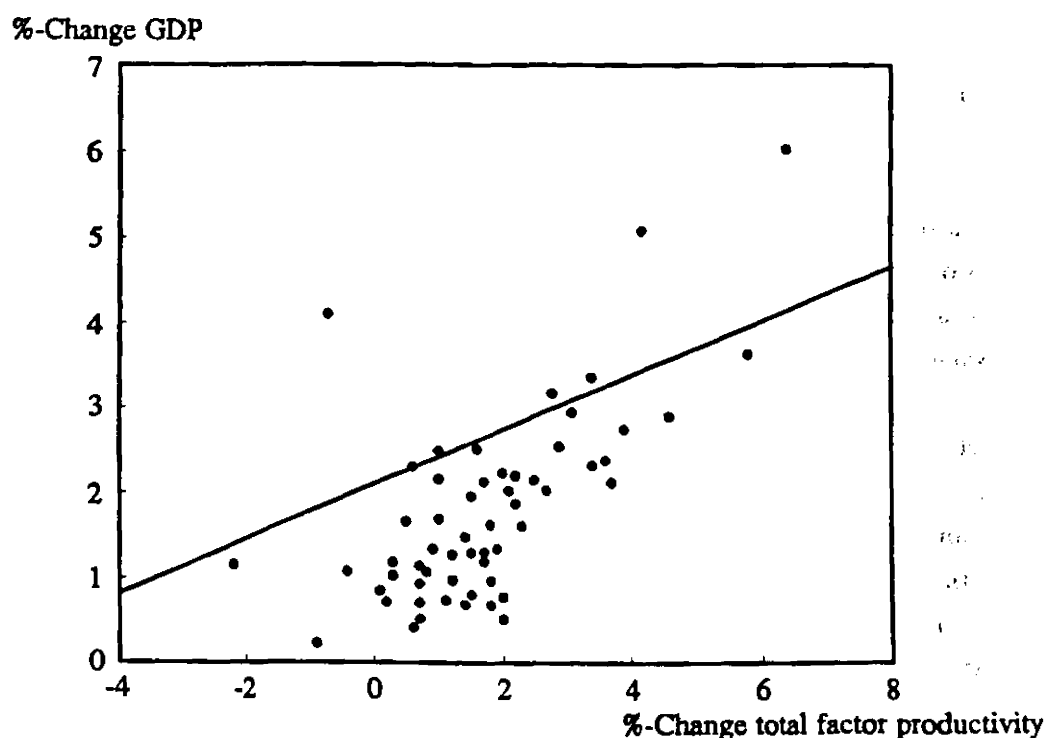
Figure 1.1 Total factor productivity growth related to percentage changes in GDP

Given the importance of technological progress for economic prosperity it is not surprising that public policy makers show great interest in the market for innovations. Exemplary is that ever since its birth the European Commission has been concerned with the functioning of this market, an interest which has been (and still is) gradually increasing over time (see Guzzetti [1995]). This concern is triggered by the belief that the market for innovations fails. Although firms have many reasons for investing in R&D, still market forces are believed to be inadequate for directing an optimal amount of funds towards R&D investments. It is the correction of this market failure with which Part I of the present study is concerned.

1.2 Private and social rates of returns to R&D: a wedge

Among the many reasons why governments should intervene in market processes, the failure of markets to deliver 'adequate' amounts of commodities (be it either too much or too little) features most prominently in economists' research agendas. Given the recently re-opened eye to the importance of technological progress for economic prosperity it is not surprising that failures on markets for innovation have received much attention in recent economic literature.[4] There it is perceived that

---

[4] It is common practise to distinguish process innovation form product innovation. The former type of R&D refers to advancements which lower the cost of production (which could lead to a reduction in price), while the latter form

private incentives to invest in R&D are too low vis-à-vis social incentives. The present section reviews the main arguments as to the existence of this difference, followed by a discussion of the relevant empirical literature.

1.2.1 The wedge created[5]

The first reason why the market for innovations fails is that private rates of return to R&D investments are lower than social return rates. This difference in rates of return is firstly due to bargaining problems. To the extent that innovative products or processes contain complementary technologies, incremental rents from an innovation may accrue to downstream suppliers. But these benefits are not (fully) counted by an upstream firm when deciding on its R&D investment. In addition to this vertical bargaining problem, a horizontal bargaining problem arises if innovative firms are not able to price discriminate perfectly in the market where they try to sell (or license) their innovation. For instance, it is difficult for possible buyers to value an innovation without the inventor giving away some information, diminishing the latter's bargaining power due to difficulties in reclaiming 'loaned' information.

Involuntary leakage of innovative information, technological spillovers, is a second source for divergence between social and private rates of return to R&D investment. This free flow of information is not counted as a benefit by the innovator, whereas a social planner takes into account all proceeds of the R&D process. Indeed, as observed by Arrow [1962], authorities are confronted with a trade-off. On the one hand, the diffusion of technological knowledge should be kept to a minimum in order to preserve private incentives to conduct R&D, but on the other hand, it is in society's interest to disseminate innovative knowledge as much as possible.

A second reason why the market for innovation fails is due to the pecuniary externality effect of technological spillovers. That is, each innovator's search for new knowledge may strengthen, through spillovers, its rival's competitive position, a socially desirable strengthening of competition which, however, induces the innovator to devote less resources to R&D.

Uncertainty surrounding the markets for innovations is a third reason for under-investment in this market. Each research project involves a great deal of trial and error (see e.g. Kamien et al.

of R&D points to the development of new products. As noted by Spence [1984, p.101]

> Sometimes the cost reducing investments operate directly on costs. In many instance, they take the form of developing new products that deliver what customers need more cheaply. Therefore product development can have the same ultimate effect as direct cost reduction. In fact if one thinks of the product as the services it delivers to the customer (in the way that Lancaster pioneered), then product development often is just cost reduction.

Indeed, most of the literature on R&D considers process innovations. Product innovations will be discussed in Section 1.5.6.

[5] This subsection is mainly based on Katz [1986], Katz and Ordover [1990], and the lectures Paul Geroski gave during the 1996 summer-school on "Empirical research in Industrial Economics", as organized by the *Center for Industrial Economics* of the *University of Copenhagen*.

[1992]), thus making it difficult to predict accurately whether or not some technology being developed will perform properly. In addition to this *technological uncertainty* is *market uncertainty*. That is, there is no perfect foresight as to the existence and size of a market for some product yet to materialize. Thirdly, *if* R&D leads to successful new processes or products, for which there is a market, it could be that some innovator's competitor has entered this market before, rewarding it with the possibility of creating entry barriers such as patenting preemptively (see Gilbert and Newberry [1982]). Indeed, any firm can lose a patent race, despite the success of its research unit, leading to *competitive uncertainty*. All in all, uncertainty in the markets for innovations restricts firms' research horizons, thereby excluding possible socially beneficial research avenues.

1.2.2 The wedge estimated

It is one thing to identify forces creating a discrepancy between private and social incentives to R&D investments, it is another to estimate this difference. Obtaining (sensitive) data on individual firm's revenues, costs and innovative activities, in order to estimate the private return rate, can be very difficult indeed. But even more troubling is the quantification of social rates of return to R&D investments. That is, it is very hard to track all social benefits originating from an innovation. And to value them is possibly even more difficult. If, for example, some new medicine saves a life, what then is the social return to the R&D investment leading to the discovery of this particular new drug? The reader, of course, can come up with as many examples as there are innovations to illustrate the difficulty, if not the impossibility, of quantifying social rates of returns to R&D investments. As Mansfield et al. [1977, p.240] remark

> The measurement of social and private rates of return from investments in new technology is an extremely difficult business, which is one good reason why so few measurements have been attempted.

But even if both return rates are estimated accurately, the difference between them only accounts for part of the divergence between social and private incentives towards R&D investment, given that there are at least two additional forces (uncertainty and pecuniary externalities) responsible for the spread in incentives.

Yet, some studies report estimates of both social and private rates of return to R&D investments. An impressive example of such an analysis is the above cited study of Mansfield et al. [1977]. They carefully track all possible social benefits arising from 4 process innovations and the introduction of 13 new products, all having occurred within some time span. The social rate of return is defined as the internal rate of return for the change in total surplus (the latter being the sum of consumers' surplus and producers' surplus), due to the innovation, calculated for that time period in which the net social benefit of the innovation is nonzero. When estimating this change in social surplus, Mansfield et al. [1977] control for costs associated with uncommercialized R&D, for parallel R&D costs (that is, costs incurred by firms pursuing the same research avenues as the eventual

innovator, but were too late with introducing the new product or process), for forgone profits due to product replacement, and for profits earned by imitators. After constructing, with equal care, private return rates, Mansfield et al. [1977] conclude that the private rate of return to R&D investment is indeed much lower than the social return rate (a maximum difference of 280 percent (!) is reported). Moreover, in about 30 percent of the cases, in retrospect firms would not have invested in the innovation, since the private rate of return was too low. But in all these cases, the social rate of return made the investment worthwhile. According to Mansfield et al. [1977], this could suggest that some socially desirable R&D projects are not undertaken. Regression results, explaining the gap between social and private return rates to R&D investments, are also presented. These estimates could be read as corroborating the importance of technological spillovers in that the gap is widened the less expensive it is to imitate an innovation, that is, the higher are technological spillovers.

The only valid critique of the Mansfield et al. [1977] study concerns the sample size. Detailed examination of both private and social return rates to R&D capital naturally limits the number of innovations to be considered. But given that only 17 innovations are examined (which, according to Mansfield et al. [1977, p.222] cannot be regarded as randomly selected), *a priori* the results should not be treated as revealing a general pattern. One way to overcome this sample size restriction is to adopt a more narrowly defined, but straightforwardly computable social rate of return. Indeed, Bernstein and Nadri [1988] define in effect the social rate of return to some innovation as the concomitant private rate of return *plus* all spillovers emanating from it. Hence, all that is needed to indicate the wedge between social and privates return rates are (accurate) estimates of technological spillovers. Observe however that this will result in conservative approximations of the wedge, because, as explained in the previous subsection, there are (at least) two other forces responsible for the gap between the different return rates.

Despite difficulties with measuring technological spillovers,[6] studies reporting estimates of these externalities abound. One strand within this literature considers *intra*industry spillovers, while another concentrates on *inter*industry spillovers. The former externality refers to the free flow of information between firms active in the same industry, whereas the latter indicates the leakage of knowledge between competitors operating in different sectors. Scherer [1982], Jaffe [1986], Bernstein and Nadri [1988], and Bernstein [1989] all report statistically significant interindustry spillovers. Bernstein and Nadri [1989], and Caballero and Jaffe [1993] on the other hand find significant intraindustry spillovers. Yet others estimate both externalities simultaneously. For example, using annual observations on 170 Canadian firms from seven two-digit Standard Industrial Classification industries[7] for a period of 13 years, Bernstein [1988] estimates both intraindustry and interindustry technological spillovers. He concludes that both types of spillovers reduce the average cost of production, although the effect of interindustry spillovers is larger than that of intraindustry spillovers.

---

[6] In particular, knowledge as such is not directly observable, technological spillovers can be confused with other measurement errors, and they are difficult to distinguish from increasing returns (see Geroski [1996]).

[7] These are Food and Beverage, Pulp and Paper, Metal Fabricating, Non-Electrical Machinery, Aircraft and Parts, Electrical Products, and Chemical Products.

Moreover, interindustry spillovers unambiguously lower the demand for R&D capital whereas intraindustry spillovers only do so in industries with a relatively low demand for R&D. For R&D intensive industries the effect of intraindustry spillovers is the reverse. An explanation for this reversal could be, as observed in the introduction, that firms have to conduct research themselves to be able to appropriate the technological spillover. Finally, Bernstein [1988] concludes that the difference between social and private rates of return to R&D are mainly due to the existence of intraindustry spillovers.[8]

All in all, technological spillovers are not merely a theoretical concept, but are also persistently revealed in empirical studies. If anything, these findings confirm the existence of a wedge between private and social rates of return to R&D investments. And this wedge only results in a conservative estimate of the difference between social and private incentives to invest in R&D given that other pecuniary externalities and uncertainty are not taken into account. Hence, leaving the R&D investment decision to the market induces firms to under-invest in R&D. To try to correct this market failure, authorities can choose one of several policy tools, or devise a comprehensive policy portfolio. What these instruments are is discussed next.

1.3 The first policy: patents

The policy spectrum for private R&D investments consists of three tools. Rewarding an innovator with some monopoly power by granting it a patent is a traditional practise.[9] This should increase the incentives to devote resources to R&D, since technological spillovers are diminished. Indeed, not only are rivals' competitive edges dulled by an innovators' research, but the latter's (horizontal) bargaining power is also strengthened, thus facilitating the appropriation of returns to R&D. Moreover, to the extent that technological spillovers enhance social welfare, the social rate of return to R&D declines, thereby narrowing the gap between social and private incentives to invest in R&D.

However, empirical research does not confirm strongly the alleged benefits of the patent system. It appears to be very difficult to prevent technological knowledge from spilling over. For example, Mansfield [1985] finds for his sample of 100 American firms, that within 12 to 18 months an innovator's decision to develop a new product or process is known to its competitors. Moreover, within one year after a new product or process is developed, its nature and operation is in rivals' hands. Indeed, as discussed in the previous section, studies reporting significant spillovers despite intellectual property rights abound. This is reflected, for example, by the observation of Mansfield et

---

[8] Observe that technological spillovers are not bounded by national borders. Indeed, Coe and Helpman [1995] estimate technological spillovers between countries. Another interesting contribution is that of Suzuki [1993], who approximates the externality as it occurs both within and between different 'keiretsu' groups.

[9] Designing a patent involves the construction of some mix between the time it lasts (the length of the patent), and the spread of adjacent innovations it precludes (the breadth of the patent). The optimal length-breadth mix is subject of an ongoing economic debate. At the heart of this discussion lies the relationship between social welfare and post innovation profits. Indeed, the literature on patents is vast and we will not try to review it here. For a recent survey see e.g. Denicolò [1996].

al. [1981], that the majority of the 48 innovations they consider would have been carried out even if there was no patent protection.[10]

But even if the patent system would work perfectly, it has its costs. Granting air-tight patents to innovators prohibits efficient sharing of innovative information, since the marginal costs of information sharing are often close to zero. Licensing contracts however, are likely to include prices faced by the licensee which are well above these marginal costs. Also, as suggested by Katz and Shapiro [1985], dummy licensing agreements involving royalty-labelled side payments could cover up cartel agreements. Moreover, as revealed in the sample of Mansfield et al. [1981], patents could be used to increase imitation costs (and with it, the time it takes to imitate) which can serve as a barrier to entry. Indeed, Mansfield et al. [1981] report regression results revealing a strong, positive correlation between imitation costs and industry concentration.

An even more fundamental critique of the patent system comes from the observation that large technological spillovers persist in R&D intensive industries (see e.g. Levin [1988]). This implies that it is questionable whether these externalities undermine firms' incentives to invest in R&D. Indeed, Cohen and Levinthal [1989] argue that R&D activities not only generate new knowledge but also enhance the ability to exploit and assimilate existing knowledge. Hence, technological spillovers may even encourage firms to invest in R&D. It is for this reason that Geroski [1996] suggests policy makers to try to maximize spillovers.

1.4 The second policy: R&D-cooperatives

A policy tool recently discovered by policy makers to fight the difference between social and private incentives to invest in R&D, is allowing firms to cooperate in their research.[11] Indeed, in the U.S., Europe and Japan, cooperatives in R&D are not illegal *per se*.[12] On the contrary, national and supra-national authorities actively encourage the formation of such strategic alliances, expecting them to gear private incentives to invest in R&D towards the socially desirable level.[13] In this section the main arguments against and in favour of this permissive antitrust treatment of cooperatives in R&D are

---

[10] This result depends very much on which industry is considered though. Especially in the pharmaceutical industry patent protection is considered to be important (see Mansfield et al. [1981]).

[11] Katz and Ordover [1990] consider licensing of innovations to be a form of *ex post* cooperation among firms. Cooperation in R&D, in which firms jointly set their R&D investments, is then considered as *ex ante* cooperation. The remainder of this chapter thence concerns *ex ante* cooperation.

[12] In Europe, the European Commission granted in its Regulation 418/85 a thirteen-year block exemption under Article 85 para.3 to collusion in R&D. In the United States some cooperation between innovating firms is allowed under the National Cooperative Research Act of 1984. Japanese corporate law also allows firms to cooperate in their research. For an elaborate comparison between Europe, the U.S. and Japan on this issue see Martin [1997].

[13] See, for example, the European Commission's "Green Paper on Innovation" (EC [1995]) and its "Action Plan for Innovation in Europe" (EC [1996b]).

reviewed.[14] The following subsection contains an overview of the formal economic literature rigorously examining the issue of cooperative R&D.

1.4.1 The case for cooperative R&D

The first main argument in favour of cooperative R&D emerges from the existence of technological spillovers. As discussed in Section 1.3.1, the mere existence of these externalities creates a wedge between social and private incentives towards R&D investments. However, technological spillovers do not diminish the incentives to invest in R&D if partners agree to share the research cost before the R&D investment is actually realized. Indeed, firms understand that there will be no free-riding on their investment, at least not by the members of the cooperative. Moreover, if future partners also agree to fully share the results generated by their cooperative R&D, the trade-off identified by Arrow [1962] (see Section 1.3.1) disappears. That is, *ex ante* agreements to fully share the fruits and cost of research preserve the private incentives to undertake R&D while diffusing all information within the cooperating group. It follows that, from this point of view, a cooperative should include as many firms as possible (within a certain industry). In addition, full sharing of information eliminates wasteful duplication within the cooperating group and thus increases the efficiency of research, that is, fewer resources are needed to obtain a given level of effective R&D. Hence, a fall in resources devoted to R&D because of cooperation does not necessarily mean a drop in effective research.

A second argument in favour of R&D-cooperatives is one of scale economies. As early as 1952, John Kenneth Galbraith observes (Galbraith [1980, p.86])

> There is no more pleasant fiction than that technological change is the product of the matchless ingenuity of the small man forced by competition to employ his wits to better his neighbor. Unhappily, it is a fiction. Technological development has long since become the preserve of the scientist and the engineer. Most of the cheap and simple inventions have, to put it bluntly and unpersuasively, been made. Not only is development now sophisticated and costly but it must be on a sufficient scale so that success and failures will in some measure average out.

He consequently concludes (Galbraith [1980, p.87])

> Because development is costly, it follows that it can be carried on only by a firm that has the resources which are associated with considerable size.

Due to the ever increasing complexity of innovations and the concomitant rise in development costs, many claim that today even large firms do not have the necessary assets for developing new technologies. R&D-cooperatives are then a natural means for participants to provide the necessary capital by pooling members' financial resources, given that capital markets are imperfect (that is, credit

---

[14] The discussion is based on Katz [1986], Jacquemin [1988], Katz and Ordover [1990], and Geroski [1993].

is rationed because of asymmetric information and different attitudes between banks and firms towards risk, the former being relatively risk averse). Moreover, to obtain financial credit the bargaining power of a consortium of firms is likely to be stronger than that of a single firm, implying that an R&D-cooperative can undertake larger projects than the sum of its members' individual capital stocks would permit. Further, what often matters for innovations to be commercially successful is the timely introduction of new products or technologies, which requires R&D capital to be readily available. Providing these assets by pooling financial resources is then the preferred alternative for the elaborate bargaining process between firms and financial institutions will take much more time than financing the project directly out of the R&D-cooperative's capital stock. In short, cooperatives can conduct larger R&D projects than any of the participants could pursue on their own, while also being able to introduce new innovations more quickly because of better access to the necessary financial resources.

Finally, the scope of feasible R&D projects is enlarged if firms conduct R&D-cooperatively. Indeed, by pooling resources more avenues of research are within reach because of possible synergies and complementarities among cooperating firms. Moreover, by sharing the risk of research through cooperation, members of the R&D-cooperative might contemplate more risky projects, thereby also widening the potential research horizon.

1.4.2 The case against cooperative R&D

Against the potential benefits of R&D-cooperatives inevitably lurk some drawbacks, the most important among these being that members of an R&D-cooperative could use the agreement as a forum to discuss and to set prizes prevailing in the product market, practices which are likely to yield a harmful reduction of competition. Indeed, not only are side payments more easily made if firms meet on a regular (and legal) basis, the fact that some firm is cooperating in the pre-competitive stage of the production process might signal its cooperative nature and it could thus be identified as a possible colluding partner in the competitive stage.[15] On the other hand, as argued by Geroski [1992], for an innovation to be commercially successful there have to be strong links between those conducting the research and those marketing the concomitant innovative product, since it is the feedback from output markets that directs research towards profitable avenues.[16] Given then the necessity of strong links between pre-competitive R&D and final competition, firms' incentives to conduct joint research might be diluted if not at least some form of joint exploitation of cooperative research is allowed for in the product market. In sum, if firms are allowed to cooperate in their research they have ample opportunity to extend the cooperative agreement to output markets (see Martin [1996] for an illustration of this point). Moreover, in order not to undermine the incentives for cooperating in R&D some market power has to be given to the cooperating firms. Indeed, as aptly observed by Jacquemin [1988], in deciding

---

[15] According to Geroski [1993], this learning about partner's behaviour is an environmental externality.

[16] For this reason, Geroski [1992] argues further that the stimulation of vertical links within innovative industries is likely to be more effective in bringing research to a success than allowing for horizontal links.

whether or not firms are allowed to form R&D-cooperatives, authorities are faced with a Schumpeterian trade-off between static and dynamic efficiency.

Another objection to R&D-cooperatives is that they might actually lower the incentives to conduct research. This is likely to happen when negative pecuniary externalities prevail. For instance, if joint research lowers only the cost of some of the participants, thereby reducing final profits of all other members of the cooperative, it could be in the interest of the whole group to diminish the overall research intensity. Another such situation arises when partners realize that products embodying new technologies will only replace existing ones. Indeed, a large dominant incumbent firm could engage in cooperative research with a small entrant in order to hamper the latter's research agenda. Also, if competition in output markets is strong, firms can jointly decide to cut on R&D expenses, since in this case severe product market competition is likely to direct all surplus towards final consumers. Finally, if future research will be such that only the first successful innovator will make a profit ('the winner takes it all'), thereby driving all other firms out of the industry, existing firms may decide collectively not to engage in this 'patent race'. What this all adds up to is that cooperatives in R&D internalizing negative pecuniary externalities are likely to collectively decide to cut on R&D activities. This then will lead to the introduction of fewer new (and possibly superior) products, not only because any cooperative agreement diminishes competition (see Martin [1994]), but those engaged in research will devote fewer resources to it.

The final main argument against R&D-cooperatives is that they can act as a barrier to entry. Indeed, excessively accelerating R&D programmes in order to patent preemptively (see Gilbert and Newberry [1982]) might very well require the amount of capital only a cooperative can come up with. Moreover, cooperating firms could collectively decide on standards of future products, effectively blocking non-participants' and future entrants' ability to compete in the post-innovation market. Industry-wide adoption of a standard set by an R&D-cooperative is, of course, more probable the more firms join the R&D-cooperative. Indeed, the threat of increased entry barriers through cooperation in R&D cautions the approval of industry-wide R&D-cooperatives.

Having discussed the main potential benefits and drawbacks of R&D-cooperatives, their functioning within economic systems is to be considered in more detail. That is, the formal economic literature on R&D-cooperatives is now examined.

1.5 The formal analysis of R&D-cooperatives; a literature review

Although the literature on strategic R&D, as it has developed over the last decade, has its roots in Katz [1986], an important catalyst to this strand of economic thought is the famous example of d'Aspremont and Jacquemin [1988].[17] They present a two-stage duopoly with homogeneous products. In the first stage both firms simultaneously determine their R&D efforts. The subsequent stage entails Cournot competition in the product market, given the R&D efforts of the first stage. Within this framework

---

[17] This paper already found its way into mainstream economics textbooks (see Shy [1995] and Phlips [1996]).

three different scenarios are considered: first, no cooperation in either the first or the second stage; second, cooperation in R&D and competition in output; and third, cooperation in both the first and the second stage. In order to assess these different market structures, the concomitant market equilibria are compared with the first best (total surplus maximizing) market solution. An important element in the analysis is the explicit modelling of technological spillovers.

D'Aspremont and Jacquemin [1988] conclude that cooperation in R&D (but not in production) leads to an increase in R&D efforts (and production) if technological spillovers are substantial (in their example this refers to technological spillovers of at least 50%). In case of small spillovers the reverse holds. Overall, the fully cooperative game leads to the highest level of R&D investments when spillovers are above 41%. As observed by d'Aspremont and Jacquemin [1988, p.1135], this is due to cooperating firms being able to capture more of the surplus generated by their R&D if competition in the product market is weakened. For technological spillovers below 41% monopoly R&D efforts are in between those generated under the fully competitive regime and those induced by the partially collusive market structure. On the other hand, production of the monopoly is always short of that under the partially cooperative and fully competitive regimes.[18] Finally, under all regimes considered and for all technological spillovers, both production and R&D efforts fall below those considered first best.

From the analysis of d'Aspremont and Jacquemin [1988] two main conclusions can be drawn. First, there is corroboration of the idea that R&D-cooperatives enhance private R&D spending if pre-cooperative technological spillovers are large (see Section 1.5.1).[19] The deeper explanation for this finding lies in the interaction between two externalities associated with strategic R&D investments. On the one hand, devoting resources to R&D activities increases the innovator's efficiency of production and thus rewards it with a larger market share at the expense of its competitors. That is, any firms' R&D investment has an impact on *all* firms' profits. On the other hand, some of the benefits of every R&D investment will spill over to rivals, thereby increasing their production efficiency. The first of these two effects, labelled by Kamien et al. [1992] as being a *combined-profits* externality, can either be positive or negative.[20] The second, identified as a *competitive-advantage*

---

[18] See also d'Aspremont and Jacquemin [1990]. Henriques [1990] points out that d'Aspremont and Jacquemin [1988] neglect to examine stability conditions. In case of small spillovers, these conditions restrict the space of some of the parameters featuring in d'Aspremont and Jacquemin [1988], but do not alter any of the results reported. Moreover, the claim of Henriques [1990, p.639] that in her numerical example of the d'Aspremont and Jacquemin [1988] model an equilibrium in the fully cooperative model does not exist if technological spillovers are above 60% is obviously wrong. Indeed, inserting the parameter values employed by Henriques [1990] into the second order condition associated with the R&D stage of the monopoly shows that this conditions holds for all technological spillovers.

[19] Hinloopen [1994] shows that in the d'Aspremont and Jacquemin [1988] firms always want to cooperate in as many stages as allowed to. He shows further that total surplus increases through cooperation in R&D (while maintaining competition in the output market) if pre-cooperative spillovers are substantial. Ziss [1994] shows that for a duopoly, in general, large spillovers are necessary but not sufficient for the latter observation to hold.

[20] Observe that the combined-profits externality is the pecuniary externality among firms (as identified by Katz [1986]) *plus* the effect of any innovator's R&D activities on own profits.

externality by Kamien et al. [1992], is unambiguously negative. In deciding how much to spend on R&D, individual firms always take account of the free-rider effect (that is, of the competitive-advantage externality). R&D-cooperatives, in addition, internalize the combined-profits externality. The first conclusion of d'Aspremont and Jacquemin [1988] then implies two things. First, if there are no spillovers (and hence, no competitive-advantage externality), the combined-profits externality is negative. Indeed, if in this case firms form an R&D-cooperative, total R&D spending will be reduced. Second, if spillovers increase, the combined-profits externality gradually becomes positive. Moreover, for large enough spillovers it outweighs the negative competitive-advantage externality (although the latter's absolute size increases with expanding spillovers), inducing cooperating firms to spend more on R&D vis-à-vis non cooperating firms.

D'Aspremont and Jacquemin's [1988] first result has a direct corollary. If for some magnitude of technological spillovers cooperative R&D exceeds noncooperative R&D, it must be the case that the combined-profits (indirect) externality reacts more to a change in technological spillovers than the competitive-advantage (direct) externality, given that in absence of the latter the former is negative. Further, observe that the two externalities are related. That is, a stronger negative competitive-advantage externality enhances the effect any innovator's R&D expenses have on its rivals' profits by widening the channel through which information flows freely. The interaction between the two externalities is illustrated in Figure 1.2, where the maximum spillover size is denoted by β_{max}, and where cooperative and noncooperative R&D investment coincide for β^* (this being 50% in d'Aspremont and Jacquemin's model [1988]).

Another way of comparing cooperative R&D with noncooperative R&D is by distinguishing complementary from substitutable research (see Geroski [1993]). As will be shown in Chapter 2, for d'Aspremont and Jacquemin's [1988] model the condition under which cooperative R&D (with competition in production) exceeds noncooperative R&D corresponds exactly to that which determines when research is complementary (see also Vonortas [1994], Poyago-Theotoky [1995] and Steurs [1995]). That is, R&D-cooperatives conduct more research than the total of research their members would undertake on their own when R&D is complementary, whereas the reverse holds for substitutable research.[21] If R&D is substitutable, negative pecuniary externalities prevail and are likely to yield an overall negative combined-profits externality, especially when the number of firms in the industry is large. Added to the free-rider effect, however small this may be, it leads cooperatives to cut their R&D expenses. Pecuniary externalities are positive however if research is complementary, implying the combined-profits externality to be positive. And this is likely to exceed the free-rider effect, given that for complementary research free riding is not a severe weakening force on individual R&D investment incentives. Hence, in this case R&D-cooperatives collectively decide to invest more in R&D.

D'Aspremont and Jacquemin's [1988] second main conclusion is that although R&D-cooperatives can increase the incentives to conduct research, they still do not result in the first-best

---

[21] d'Aspremont and Jacquemin [1988] do not report this equivalence.
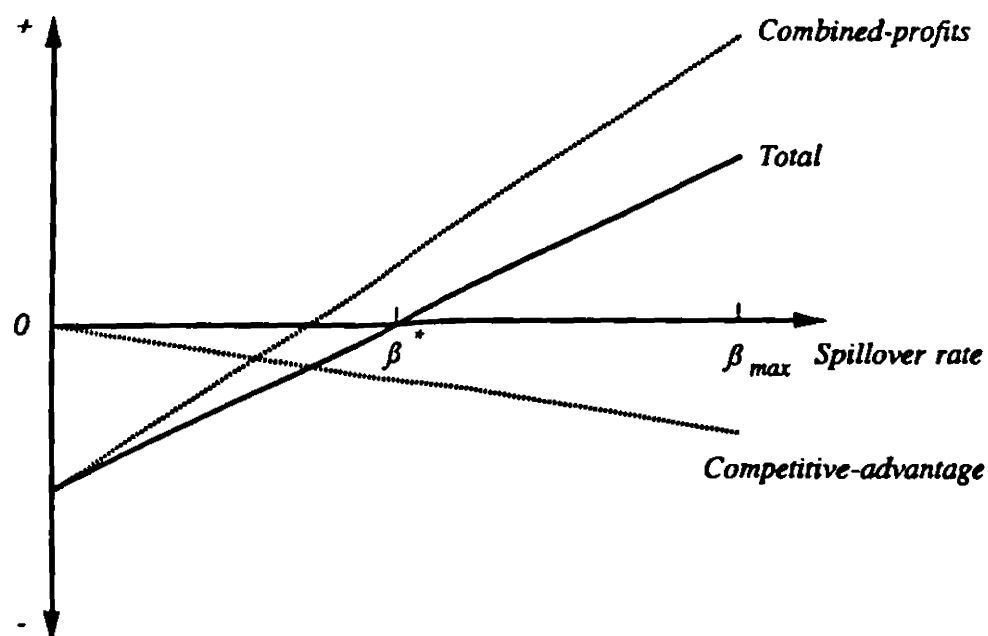
Figure 1.2 R&D externalities as a function of technological spillovers

level of R&D investment.[22] Moreover, if pre-cooperative spillovers are modest, allowing firms to cooperate in their research will lead to a reduction in total resources being devoted to R&D, inducing an even larger gap between the first best R&D investment and that provided by the market. If anything, the seminal analysis of d'Aspremont and Jacquemin [1988] suggests caution towards a too permissive antitrust treatment of R&D-cooperatives.

Given the elegant yet simple set-up in d'Aspremont and Jacquemin [1988], there is ample scope for generalization. Indeed, since 1988 many studies provided such generalizations in one direction or another. In what follows some of these will be discussed in detail. First however a closer look is to be taken at what is meant exactly by cooperating in R&D, because an implicit assumption in d'Aspremont and Jacquemin [1988] as to the relation between pre-cooperative technological spillovers and R&D-cooperatives, caused the subsequent literature to label different types of R&D-cooperatives chaotically.

---

[22] This result appears to hold for a wide class of models, as shown by Suzumura [1992] (his analysis will be discussed in extensive detail in Chapter 3). He also shows that cooperative R&D falls below some second-best level of R&D investments, that being defined as those R&D efforts a social planner would set *given* the product market outcome. *Non*cooperative R&D can, however, exceed this level if the number of firms in the industry is large enough (this is also reported by Yi [1996]). In Chapter 2 these results are extensively elaborated upon, since it has a drastic effect on the provision of R&D-subsidies.

1.5.1 On the definition of R&D-cooperatives

Cooperation among firms in one or some stages of the production process can range from an occasional sharing of information to maximization of joint profits. But the literature on cooperatives in R&D is not clear as to the labelling of these agreements. On the one hand different names circulate to identify the same type of R&D-cooperative, whereas on the other hand the same agreement for cooperation in R&D is named differently in different studies.

Kamien et al. [1992] order this lexicographic jungle by making explicit the difference between R&D-cooperatives in which participants agree to share information as to their individual research outcomes, and agreements according to which firms jointly set their R&D investments without exchanging innovative information. Indeed, this clarification is Kamien et al.'s [1992] reaction to d'Aspremont and Jacquemin's [1988] implicit assumption that an agreement to cooperate in R&D does not affect the extent to which information flows among members of the cooperative.[23] To avoid confusion as to what is meant in the present study by cooperation in R&D the (implicit) definitions given by Kamien et al. [1992] are adopted.[24] These read as follows.

Definition 1.1 R&D-cartel

An agreement to coordinate R&D activities so as to maximize the sum of overall profits

This contract does not imply that participating firms share the outcomes of their R&D efforts, that is, pre-cooperative technological spillovers are not affected by firms participating in this cooperative agreement. Rather, firms form a strategic alliance and set individual R&D investments such that joint profits are maximized. It is this type of R&D-cooperative d'Aspremont and Jacquemin [1988] analyze.

Another type of agreement emerges if competitors only agree to exchange innovative information.

Definition 1.2 Research Joint Venture (RJV)

An agreement in which firms decide unilaterally on their R&D investments but the results of their R&D are fully shared

---

[23] Two years later though the same authors do indicate that cooperation in R&D will lead to higher spillover rates (see d'Aspremont and Jacquemin [1990]). Moreover, the results reported by d'Aspremont and Jacquemin [1988] do not change qualitatively if the fixed spillover assumption is relaxed. Indeed, as shown by Choi [1993], allowing the spillover rate to increase with cooperation in R&D still induces the R&D-cooperative to invest too little in R&D vis-à-vis some second-best level. Moreover, also in this case the R&D-cooperative only devotes more resources to R&D than their competitive counterpart if pre-cooperative technological spillovers are large.

[24] Kamien et. al [1992] do not state exactly the same definitions as those presented in the text. However, they can be unequivocally deduced from their descriptions of the different types of R&D-cooperatives.

Note that RJVs do not entail cooperation in R&D to the extent that R&D investments are coordinated. Rather they indicate the sharing of results of *independent* R&D efforts. Indeed, within RJVs duplication of research is diminished and possibly eliminated. The extent to which duplication is avoided depends on the frequency with which members of an RJV exchange their research results, whether or not partners fully disclose their R&D records, and the type of research firms cooperate in (generic research being more transparent and easy to diffuse than development research).

The final agreement to cooperate in R&D, as distinguished by Kamien et al. [1992], involves both notions of cooperation.

Definition 1.3   Research Joint Venture Cartel (RJV-cartel)

An agreement to fully share the results of R&D and to coordinate R&D activities in order to maximize the sum of overall profits

Indeed, as to the intensity of collaboration between firms, an RJV-cartel entails the furthest reaching type of cooperative agreement. Observe that both in an RJV or RJV-cartel firms could be seen as licensing their innovations at zero price.

1.5.2 RJVs, R&D-cartels, and RJV-cartels

Analyzing the scenarios of Section 1.5.1 within an extended version of d'Aspremont and Jacquemin's [1988] model,[25] Kamien et al. [1992] conclude that resources devoted to R&D in an R&D-cartel exceed the sum of competitive research if technological spillovers are 'high enough', while the reverse holds for 'small' pre-cooperative technological spillovers, thereby confirming the first main result of d'Aspremont and Jacquemin [1988]. The dividing line between 'high' and 'low' now depends however on the extent to which products are differentiated. If products are homogenous, 'high' means above 50% (as in d'Aspremont and Jacquemin [1988]). But this threshold (linearly) diminishes with the extent to which products are differentiated, and vanishes if products are independent (that is, being differentiated to a maximum). In terms of Figure 1.2 it means that β^* moves to the left the more differentiated products are. The mechanism moving this threshold consists of two influences. First, the more distinct products are the less a free flow of innovative information dilutes competitors' incentives to invest in R&D. Second, competitors' profits are lowered less by an innovators' success the more differentiated their products are from that of the innovator. Moreover, it pays a firm more to invest in R&D the more monopoly power it has in the concomitant product market, that is, the more distinct its products are from all other commodities supplied. All in all, the combined-profits externality is more likely to overrule the competitive-advantage externality the more products are differentiated. In

---

[25] Kamien et al. [1992] consider an industry comprising *n* firms, allow products to be differentiated and hence consider also Bertrand competition in the product market, and treat the magnitude in marginal cost reduction through R&D efforts as being given by a general R&D production function.

the limit, when products are independent, the combined-profits externality is always positive, since in this case any firms' R&D investment only affects (positively) its own profits. And given that this externality reacts more to a change in technological spillovers than the competitive-advantage externality (see Section 1.5.1), it follows that for independent products an R&D-cartel devotes more resources to R&D than with competitive R&D.

The second conclusion of Kamien et al. [1992] is that for all intensities of technological spillover industry-wide RJV-cartels yield the highest level of R&D investment, while the formation of industry-wide RJVs induces the amount of resources devoted to R&D to be the lowest among all scenarios considered. Indeed, in an RJV the competitive-advantage externality is at its maximum while not being tempered by the combined-profits externality. On the other hand, the diverting effect of technological spillovers between the indirect and direct externality is at its maximum when information is fully shared (see Figure 1.2). Hence the maximum R&D investment is in an RJV-cartel.

Kamien et al.'s [1992] main conclusion is that the formation of an industry-wide RJV-cartel yields both the highest consumers' *and* producers' surplus. This means that the formation of this type of R&D-cooperative is socially desirable. Moreover, no government intervention in the market process is required for these R&D-cooperatives to emerge.

This analysis is taken one step further by Kamien and Zang [1993]. They use the model developed in Kamien et al. [1992] to analyze a market in which there are several competing RJV-cartels of equal size. Indeed, as suggested by Katz [1986] and Jacquemin [1988], any R&D-cooperative should be accompanied by a competitive fringe or another R&D-cooperative in the same industry to avoid concentration of market power. Kamien and Zang [1993] allow for spillovers between RJV-cartels. That is, in contrast to Katz [1986], RJV-cartel members are not assumed *a priori* to dampen completely the leakage of technological information from the RJV-cartel to the outside world.

Under certain conditions on the number of firms per RJV-cartel, the degree of substitutability of final products, and the between RJV-cartel spillovers, Kamien and Zang [1993] show that an industry hosting several competing RJV-cartels will conduct more R&D and quote lower product prices than would an industry-wide RJV-cartel.[26] Moreover, among all possible splittings of the industry-wide RJV-cartel, dividing it in two competing RJV-cartels always yields the lowest product prices. These findings are explained by identifying three factors. First, the competitive force working among competing RJV-cartels, inducing cooperatives to invest in R&D, declines as the number of RJV-cartels falls. Second, each RJV-cartel will have more members if the total number of R&D-cooperatives declines in an industry with a fixed number of firms. This strengthens each individual RJV-cartel, inducing each to invest more in R&D. Finally, if RJV-cartel membership increases, R&D conducted within the cartel will be more efficient since more firms profit from it, thereby also increasing R&D conducting incentives. The interaction between these three forces shapes the conditions under which it is profitable to split an industry-wide RJV-

---

[26] Observe that there is an obvious logical error in Kamien and Zang's [1993] second proposition. Indeed, the second sentence should read: 'The equilibrium prices obtained by the grand RJV cartel *are lower than* that of the split industry if and only if eq. (22) is reversed'.

cartel into two competing RJV-cartels. However, Kamien and Zang [1993] also show that firms' profit always falls as a result of dividing an industry-wide RJV-cartel into any number of competing RJV-cartels. Hence, however beneficial such a division for consumers' surplus might be, cooperation from members of the industry-wide RJV-cartel is not to be expected.

The analysis of Kamien and Zang [1993] opens up an interesting route further to be investigated. In particular, their assumption as to the fixed number of firms in the industry should be relaxed. In addition, the formation of an R&D-cooperative could be treated as an endogenous process, not necessarily involving all firms joining the R&D-cooperative. A first attempt in that direction is next to be discussed.

1.5.3 Endogenous RJV-cartels

Using a member of the family of models considered by Kamien et al. [1992], Poyago-Theotoky [1995] considers the endogenous formation of one, non industry-wide RJV-cartel.[27] Her first result is that independent of the size of pre-cooperative technological spillovers, which is assumed to be the same between members and non-members of the RJV-cartel and among non-participants, firms participating in the RJV-cartel always devote more resources to R&D than non-members. This result follows readily from the discussion of Kamien et al. [1992]. Indeed, if a firm joins the RJV-cartel, part of the incentive diluting free-rider effect is completely internalized with respect to all other firms in the RJV-cartel. This encourages members' R&D efforts compared to those experiencing the 'full' free-rider effect, that is, firms not participating in the R&D-cooperative. Referring to numerical simulations, Poyago-Theotoky reports further that the first-best RJV-cartel size, being industry-wide, never emerges if the formation of the R&D-cooperative is left to the market. Hence, Poyago-Theotoky [1995] concludes that there is room for an industrial policy encouraging industry-wide RJV-cartels.

The results of Poyago-Theotoky [1995] are subject to two considerations. First, she defines the number of RJV-cartel participants set by the social planner as evolving from maximizing total industry profits. Leaving out consumers' surplus in the social planner's considerations is justified by claiming that (Poyago-Theotoky [1995, fn.15])

> Furthermore, it is not clear that consumer surplus issues are at the heart of the discussion of technology policy.

However, it is with implicit reference to an improvement in consumers' surplus that for example the block exemption on R&D-cooperatives is introduced by the European Commission in its Article 85 of the Treaty of Rome. An alternative to Poyago-Theotoky's [1995] first best criterium could be

---

[27] RJV-cartels are labelled RJVs by Poyago-Theotoky [1995].

defined as the sum of consumers' and producers' surplus.[28] Second, in defining the equilibrium size of an RJV-cartel, Poyago-Theotoky [1995] considers internal stability of the cartel (see Martin [1993, Chapter 5]) while endowing RJV-cartel members with the ability to block possible additional members. The latter ability replaces the condition for external cartel stability and can lead to situations in which firms would want to join the RJV-cartel, but are not allowed to since that will lower the R&D-cooperative's profit. Indeed, this could well be the reason why the emerging RJV-cartel is non industry-wide.

1.5.4 Generic research and development research

Katz [1986] shows that for some parameter configurations of his model the formation of R&D-cartels leads to increased levels of effective R&D. The parameter spaces for which this holds are larger the higher pre-cooperative spillovers are, or the more information can be shared effectively among participants of the R&D-cooperative. Also, in Katz's [1986] model, members of an R&D-cartel always want to exchange the innovative information generated in the R&D-cartel among themselves as much as possible. Katz [1986] thus concludes that R&D-cooperatives are more likely to increase effective R&D if they involve generic (basic) research rather than development (applied) research, since the former is associated with high spillovers whereas the latter frequently involves firm-specific (low spillover) research. Moreover, the ceiling to effective information sharing within an R&D-cartel is higher in generic research than in development research since generic research is more likely to be equally useful to all participating firms in the R&D-cooperative. In addition, as Geroski [1993] points out, development costs are much higher than costs of basic research, and reactions from output markets (which may temper the incentives to conduct R&D if negative pecuniary externalities prevail) to generic research are much weaker than (negative) feedbacks to development research.

However, as e.g. Vonortas [1994] observes, most firms operating in R&D intensive industries conduct both generic and development research. Hence, an R&D-cooperative could involve both types of innovative activity. Bozeman et al. [1986] provide a graphical argument that shows that cooperatives of this kind will conduct more generic research and less development research compared to noncooperative R&D. Their reasoning is as follows. There are two identical firms, A and B, which can invest both in generic and development research. The transformation curves between these two phases of R&D are depicted in Figure 1.3 by lines $A^D A^G$ and $B^D B^G$ for firm A and B respectively.[29] Respective isoprofit curves are Π_A^{NC} and Π_B^{NC} (where NC stands for noncooperative). It follows that

---

[28] In Poyago-Theotoky's [1995] model the condition that the sum of a single RJV-cartel member's marginal cost of production and that of a single non-participant is below the double of a single firm's marginal production cost if no RJV-cartel is formed, implies that cooperation in R&D always leads to an increase in consumers' surplus. In all simulations performed by Poyago-Theotoky [1995] this condition is met, which justifies the social planner only to consider total industry profits.

[29] The fact that firm B has larger initial endowments is assumed for expositional convenience but is not necessary for the argument to go through.
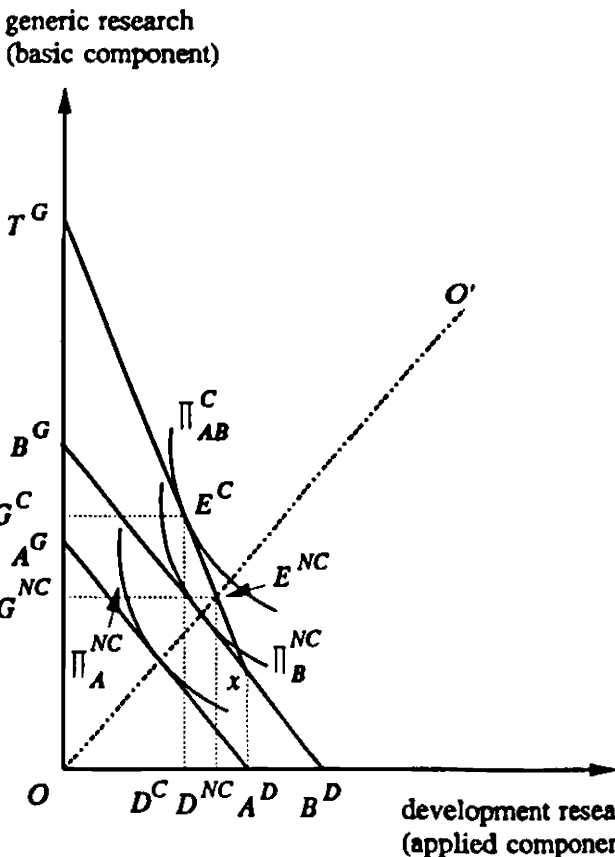
Figure 1.3 Increase in generic research through cooperation

the ray *OO'* represents the locus of optimal investment choices for the two firms if they act noncooperatively. Knowledge generated through generic research is assumed to be perfectly disseminated between the two firms. Hence, if firm *A* invests OA^G in generic research, and *B* devotes OB^G to the basic component, both firms have access to OT^G of generic research outcomes (observe that $OT^G = OA^G + OB^G$). This also implies that, in equilibrium, both firms always use the same amount of generic research and hence, by definition, the same amount of development research. A noncooperative equilibrium is thus presented by a single point on the ray *OO'*, identifying both firms' investments. Recall that OT^G is the allocation both firms can obtain simultaneously if all resources are devoted to the basic component. But there are of course an infinite number of allocations the two firms can obtain simultaneously. In particular, if firm *A* devotes all its resources to development research, and firm *B* matches this with an equal investment, both firms spend OA^D on development research. And if firm *B* spends the remainder of its budget ($B^D - A^D$) on generic research, *both* firms end up in point *x*. It follows that all locations along the line xT^G are feasible combinations of generic and

development research both firms simultaneously can obtain. Given that any noncooperative equilibrium must be on OO', it follows that E^{NC} is the noncooperative equilibrium. However, this is not a tangency point of the 'total' transformation curve (xT^G) with an isoprofit function, implying that firms can improve their performance by coordinating their actions. Indeed, tangency occurs at E^C, an equilibrium point that can be reached only through cooperation. Comparing then E^{NC} with E^C reveals that the latter equilibrium entails more generic (G^C) and less development (D^C) research than the former, respectively being given by G^{NC} and D^{NC}.

A first step towards an analytical model with which cooperatives in generic and development research can be evaluated is made by Vonortas [1994], who also builds on the framework of d'Aspremont and Jacquemin [1988]. Vonortas [1994] considers generic research to be highly in-appropriable, implying that substantial technological spillovers accompany this research phase. Development research is treated as generating firm-specific knowledge and Vonortas [1994] assumes that in this stage of the research process technological spillovers are absent. The two phases of the R&D process are assumed to be complementary, and each firm is believed to be involved in both kinds of research. Within this setting Vonortas [1994] examines the effects on market performance of allowing firms to cooperate in their basic research by either forming an R&D-cartel or an RJV-cartel.[30] In order to do so, the equilibria under the first two regimes, as distinguished by d'Aspremont and Jacquemin [1988], are computed. However, given that R&D is now treated as a two-stage process, Vonortas [1994] has to solve a three-stage game rather than a two-stage game. In the first stage, each firm determines (either cooperatively or noncooperatively) its investments in generic R&D. Given these, development research levels are determined (noncooperatively) in the subsequent stage. The final stage involves the two firms competing over quantities, given the predetermined levels of generic and development research.

The first result of Vonortas [1994] corroborates the analysis of d'Aspremont and Jacquemin [1988]: an R&D-cartel (in generic research) leads to higher levels of generic research (plus development research and production) if this research is complementary, while it reduces the incentives to invest in generic research if it is substitutable. Vonortas [1994] explains this finding by observing that if generic research is complementary, a unilateral increase by firm i in its generic research raises firm j's generic research investments. And because generic and development research are complemen-tary, rivals' total R&D efforts increase. Hence, the incentives to invest in generic research are higher if firms determine this cooperatively. On the other hand, if generic research is substitutable each firm has a (strong) incentive to invest in generic research since this will lower its rival's efforts. An R&D-cartel is then a vehicle to eliminate this strategic 'overinvestment' in generic research, which thus reduces total resources devoted to generic research.

---

[30] If the duopolists jointly maximize profits over generic research (conditional on development research and production levels), they are said to form a 'secretariat RJV' by Vonortas [1994]. If firms also agree to fully exchange the outcomes of their generic research, Vonortas [1994] labels the agreement as being an 'operating entity RJV'. Clearly, following the definitions of Section 1.5.1, the former cooperative is an R&D-cartel (in generic research), while the latter is an RJV-cartel (in generic research).

As to the comparison of RJV-cartels in generic research with noncooperative research, the results presented by Vonortas [1994] are less clear cut. For a large set of parameter configurations,[31] it appears that development research and concomitant production are higher if preceded by an RJV-cartel (in generic research) than if preceded by noncooperative generic research. On the other hand, provided that a technical condition on some parameters of Vonortas' [1994] model holds, competitive generic and development research as well as production exceed those associated with an RJV-cartel in generic research if technological spillovers in the generic research phase are below 33%.

Finally, Vonortas [1994] presents simulations results which indicate that an R&D-cartel (in generic research) is socially desirable whenever it leads to an increase in generic research (that is, when generic research is complementary), while an RJV-cartel is, from a social welfare point of view, always preferred over noncooperative generic R&D. And the latter type of cooperative agreement could be implemented without objections from the industry since, as the simulation results of Vonortas [1994] show, it always pays a firm to enter an RJV-cartel in generic research.

To conclude this subsection, among economists there seems to be agreement as to the desirability of having firms to cooperate in generic research rather than in development research, the main reason for this being that especially in generic research substantial technological spillovers persist. Moreover, as noted by Geroski [1993], generic research is apt to be much more expensive than development research. However, firms active in R&D intensive industries are most likely to be involved in both phases of the R&D process and thus could cooperate in both stages simultaneously. To date this scenario has not received proper analytical treatment. Indeed, the graphical exposition of Bozeman et al. [1986] deserves a formal analysis, possibly along the lines set forward by Vonortas [1994].

1.5.5 Intra-industry and inter-industry R&D spillovers

If anything, the work of Bernstein [1988] (see Section 1.3.1) emphasizes the importance of distinguishing between intra- and inter-industry technological spillovers. Steurs [1995] observes that no theoretical model has been developed in which these two types of externalities are formally distinguished. Accordingly, he generalizes the framework of d'Aspremont and Jacquemin [1988] by including another sector, which is the mirror of the duopoly developed by d'Aspremont and Jacquemin [1988]. The two sectors are completely independent but for the existence of inter-industry spillovers.[32] Within this framework Steurs [1995] analyzes the equilibria as they emerge under competitive R&D, and under the formation of R&D-cartels, either within or between industries.

---

[31] See Vonortas [1994, Proposition 2 and Lemma 1] for technical details.

[32] Steurs [1995] acknowledges (Steurs [1995, p.269])

> This is a strong assumption since industries between which inter-industry R&D spillovers exist are probably related in one way or another, either horizontally or vertically.

He justifies this assumption however by claiming that the introduction of dependencies between the different markets considered implies that intra- and inter-industry R&D spillovers become less distinguishable.

The first conclusion of Steurs [1995] is that if firms compete in the first stage of the game, intra-industry spillovers diminish the incentive to conduct private R&D, while inter-industry spillovers encourage private R&D efforts. That intra-industry spillovers have a diluting effect on private R&D efforts is, of course, due to the possible strengthening of competitors which outweighs the 'free' cost-reduction, as observed by Katz [1986]. On the other hand, inter-industry spillovers stimulate R&D because the two markets are assumed to be independent. That is, this externality only results in a costless reduction of marginal cost. However, in Steurs' [1995] generalization of d'Aspremont and Jacquemin [1988], inter-industry spillovers reinforce the diminishing effect on R&D investment incentives of intra-industry spillovers. Hence, the total effect of inter-industry spillovers on competitive R&D depends on the relative strength of their direct positive effect and indirect negative effect.

Referring to simulation results, Steurs [1995] also shows that the gap between the first best (total welfare maximizing) level of R&D and that emerging under a competitive market system is increasing in both spillover rates. For intra-industry spillovers this result follows readily since these spillovers are welfare enhancing while they lower private incentives to invest in R&D. Inter-industry spillovers on the other hand (directly) stimulate both competitive R&D as well as that set by a social planner. Yet the gap between these two levels widens as the inter-industry spillovers increase, since this increase amplifies the negative externality of intra-industry R&D spillovers.

Finally, comparing R&D-cartels within and between industries[33], Steurs [1995] concludes that intra-industry cooperation in R&D only encourages private R&D efforts if research is complementary, a result which is *independent* of the presence of inter-industry spillovers, while inter-industry cooperation in R&D always enhances private R&D spending if (only marginal) inter-industry spillovers persist. Moreover, in most cases an R&D-cartel between industries leads to more resources being devoted to R&D than an R&D-cartel within an industry would do.

1.5.6 Product innovation

Studies considering strategic R&D most often treat innovations as being process innovating. Indeed, as Beath et al. [1989] point out, tournament models are *a priori* more suited for describing product innovating R&D. An exception is Motta [1992], who analysis strategic R&D that increases the quality of products. Motta [1992] carefully constructs an oligopolistic model in which firms first decide whether or not to enter the industry, then either cooperatively (industry-wide) or noncooperatively set their quality improving R&D levels, and finally compete over quantities in the output market. Following Katz [1986], Motta [1992] distinguishes between the spillover rate within the cooperating group (referred to as the 'information exchange parameter'), and that between non-cooperating firms (labelled as the 'technological leakage parameter'), where the former is at least as large as the latter. The cooperative agreement does not explicitly involve any change in the information parameter. Hence, Motta [1992] considers an industry-wide R&D-cartel for product innovating industries.

---

[33] Scenarios in which there is only one R&D-cartel, either between or within industries, are not examined.

The first result of Motta [1992] is in agreement with the literature on process innovating R&D: quality levels and social welfare increase through the formation of an industry-wide R&D-cartel if technological leakage is 'high enough'.[34] The main result of Motta [1992] however is that if the leakage of information is modest (below 57%) the number of firms entering the industry is *higher* under the formation of an industry-wide R&D-cartel than under competitive R&D. In that case, the formation of an R&D-cooperative is socially desirable, since with an R&D-cartel more firms will compete in the product market, leading to lower product prices. Observe that this result is in striking contrast with the threat of an increased entry barrier due to the formation of R&D-cooperatives.

Rosenkranz [1995] also analyzes cooperative R&D agreements if successful R&D leads to product innovations. She considers a duopoly which combines elements from tournament and non-tournament models. In particular, in the first stage of the game firms decide, either cooperatively or noncooperatively, on *both* the quality and the timing of their innovation. A second stage then follows in which firms compete over price. R&D is assumed to be more expensive the higher the quality of the final product is to be, and the sooner the market is to be entered. These assumptions are justified by observing that higher quality products require more sophisticated (and hence, more expensive) research, while bringing forward the entry date requires firms to undertake parallel research projects, impeding them to learn through sequential research.

For this setting Rosenkranz [1995] proves that, depending on consumer's preferences, two equilibria exist in which one firm produces the highest quality product possible while the other settles for products of minimum quality (this maximum differentiation result in the quality domain is first reported by Gabszewicz and Thisse [1986]). The first equilibrium involves the high quality producer entering the market first, followed by the low quality producing firm, and can be sustained under a uniform distribution of consumers' preferences for quality over the unit interval. In this case, the high quality firm earns monopoly profits during the time it is the only producer. These profits are needed to cover the high R&D costs incurred from early and high quality production. A low quality producer on the other hand earns little profits, because in Rosenkranz's [1995] model consumers' willingness to pay for higher quality only depends on their initial position in the quality dimension.[35] This forces the low quality producer to wait when entering the market, not only because R&D cost are higher the sooner the market is entered, but also because it can learn from the incumbent's R&D activities through technological spillovers.[36] A second equilibrium exists in which the low quality producer

---

[34] It remains to be analyzed whether Motta's [1992] condition indicating when technological leakage is high enough, coincides with that establishing when product innovating research is complementary.

[35] Rosenkranz [1995] assumes the market to be completely covered. That is, even the consumer with the lowest preference for quality buys the low quality product.

[36] Although Rosenkranz [1995] implicitly refers to the existence of technological spillovers, they are not modelled explicitly. Her definition of an RJV-cartel (labelled as an RJV) involves information sharing among firms to the extent that as soon as the high quality firm starts producing, it will inform its partner how to produce the low quality commodity.

enters the market *first*. This can occur if consumers' preferences for quality are low.[37] In this case there is little scope for the high quality producer to differentiate its product from the lower quality commodity. Profits of the firm producing low quality output are now higher compared to the situation in which consumers' prefer higher quality. In addition, it can earn monopoly profits when entering the market first. At some point these two effects are strong enough to allow the low quality producer to enter first. This happens then when consumers' taste is biased substantially towards inferior products.

Examining what happens if firms are allowed to coordinate their research activities, Rosenkranz [1995] reports the remarkable result that then a monopoly, producing the high quality product, will emerge. She further shows that from society's point of view this is never desirable. The intuition behind the cooperative market equilibrium is that if firms are allowed to coordinate their actions in the first stage of the game, it always pays the low quality firm not to enter the market and to receive half of the high quality producing firm's monopoly profits. That is, evenly split monopoly profits exceed average profits if one firm produces a high quality product while the other supplies commodities of low quality. It should be noted however that legislation allowing for cooperation among firms in their R&D is intended for existing markets, that is, for markets in which future partners are already active. Rosenkranz's [1995] cooperative equilibrium therefore describes a hypothetical scenario.

Nevertheless, it is true that the literature examining cooperatives in strategic R&D almost exclusively deals with process innovations. The results of Motta [1992] and Rosenkranz [1995] suggest that different conclusions apply to markets in which R&D is product innovating. It is important to examine whether the result of Motta [1992], that an R&D-cooperative can sustain more firms in the market, and that of Rosenkranz [1995], that firms have an incentive to leave the market if they are allowed to coordinate their R&D, carry over to other, possibly more general models of R&D, in particular to those treating the R&D process as process innovating.

1.6 The third policy: R&D-subsidies

The theoretical literature on R&D-subsidies is scarce. The reason for this apparent gap in economic literature might be the negative connotation any subsidy-scheme evokes with academic economists. And there are indeed drawbacks to R&D-subsidies. As Katz [1986] observes, as such R&D-subsidies do not increase the dissemination of research, firms may deceive authorities to obtain the subsidy (e.g. labelling too many personnel as researchers), the market may even be further distorted possibly leading to socially excessive R&D investments, and taxes to raise the R&D-subsidy incur deadweight losses and might be politically unfeasible. Moreover, ex ante it is not clear that authorities will only subsidize successful research projects.

---

[37] See Rosenkranz [1995] for the precise meaning of 'low'.

Yet, policy makers do provide R&D-subsidies on a large scale.[38] And there are of course reasons for this. Not only are the incentives to conduct R&D enhanced through the provision of R&D-subsides (Spence [1984], Fölster [1995]), entry barriers are also lowered, implying more competition in the research market which possibly leads to the introduction of more, new and superior technologies (Martin [1995]).

Although R&D-subsidies are not popular in economics, they have received some attention in the literature. Reviewing this scarce literature is the subject of the present section. Contrary to popular belief, the main conclusion that can be drawn from this modest review is that there is a strong case to be made for the provision of R&D-subsidies. It should be noted however that all contributions discussed here ignore many of the drawbacks associated with providing R&D-subsidies mentioned above.

1.6.1 R&D-subsidies in national markets

Market performance and R&D-subsidies

One of the earliest theoretical contributions examining the effect of R&D-subsidies on industry performance is Spence [1984]. He considers a homogeneous oligopoly in which firms accumulate knowledge through research activities. Technological spillovers are explicitly taken into account, implying that each firms' R&D efforts contribute to the stock of knowledge of any other firm in the industry. The process of R&D is assumed to exhibit constant returns to scale,[39] and, as usual, it is considered to be process innovating. In this setting an active government is introduced that can provide R&D-subsidies such that *net* social welfare is maximized.

Examining a specific example in which firms compete over quantities, Spence [1984] finds that R&D-subsidies increase the incentives to conduct R&D, and improve industry performance (defined as the ratio of actual total surplus to first-best social surplus). The effect of R&D-subsidies is larger, the higher are technological spillovers. The reason for this is twofold. On the one hand, the larger are technological spillovers, the lower will be private incentives to devote resources to R&D. On the other hand, social welfare is increasing in the spillover rate, since any cost reduction is welfare enhancing. Optimal R&D-subsidies, which are determined as to maximize some kind of second-best level of net social welfare, bridge the gap between private and social incentives to conduct R&D. And this gap is broader the larger technological spillovers are.

---

[38] See e.g. European Commission [1995].

[39] According to Spence [1984] this is a harmless assumption (Spence [1984, p.104, fn.7], words in parenthesis added)

> The diminishing returns case is more realistic. But the qualitative properties of the static model here and the dynamic (diminishing-returns) model are the same.

Subsidizing R&D when racing for a patent

Using a variant of Dasgupta and Stiglitz's [1980] patent race model, in which either a monopolist innovates or in which there is a competitive research market, Romano [1989] examines the issue of subsidizing R&D. In particular, a per-research path subsidy is considered. This R&D-subsidy is derived from maximizing a second-best welfare function, in which the social cost of raising the necessary revenue to finance the R&D-subsidy is explicitly taken into account.[40]

Romano [1989] concludes that no matter how high the excess burden associated with subsidization, a monopolist should always be subsidized if the patent life is finite (in case of an infinite patent life the second-best R&D efforts and those induced by a monopolist coincide, as shown by Dasgupta and Stiglitz [1980]).[41] On the other hand, in the case of a competitive research market, optimal R&D-subsidies are only positive if the patent life is 'short' and if the social cost of subsidization are low.[42] Finally, the provision of optimal R&D-subsidies to a monopolist raises social welfare more than subsidizing a competitive research market.

1.6.2 R&D-subsidies in international markets

The majority of the attention paid to R&D-subsidies is within the literature on international trade. In this subsection the main contributions to this subset of the international trade literature are reviewed.

Deterministic R&D

In a seminal paper, Spencer and Brander [1983] examine R&D-subsidies within an international context. A duopoly with homogeneous products is analyzed in which each firm is located in a different country. R&D is treated as being deterministic and process innovating, implying that R&D efforts always diminish marginal cost of production. There are diminishing returns to the process of R&D but no technological spillovers. Firms' market behaviour is described by a two-stage game. In the first stage both firms simultaneously determine their R&D investment. Given these investments, a second stage follows in which competition is over quantities produced.[43] An active government is introduced

---

[40] Romano [1989] does not specify however if either producers or consumers are taxed for the R&D-subsidy.

[41] Initially a monopolist can be induced to increase its R&D efforts at an infinite rate through the provision of a marginally positive subsidy. And because initially the increase in R&D efforts is of a higher order than the concomitant subsidy, some subsidization is always desirable since more R&D is welfare enhancing. Hence, also under 'very high' excess burdens the optimal R&D-subsidy is positive.

[42] Romano [1989] does not allow for an R&D tax since this lacks political appeal.

[43] As to the restriction to Cournot competition in the second stage, Spencer and Brander [1983] remark (Spencer and Brander [1983, p.717-8])

> Similarly, one could modify the precise nature of final stage competition. In particular moving to price-Nash rather than quantity-Nash does not change the nature of the results, provided products are slightly differentiated.

which can credibly commit itself to the provision of an R&D-subsidy or export subsidy, prior to the R&D stage. A crucial assumption is that all production is for export to a third country. Hence, for each country net social welfare is just the domestic firms' profits plus the net balance of payments.[44] The objective of the government is to maximize social welfare by providing either an R&D-subsidy or both an R&D-subsidy and an export subsidy.

For this setting, Spencer and Brander [1983] show first that a domestic R&D-subsidy stimulates domestic R&D, if only one country provides the R&D-subsidy. Moreover, this domestic subsidy increases foreign R&D if research is complementary, while it decreases foreign R&D in case of substitutable research. Second, the optimal R&D-subsidy is positive if R&D is substitutable, whereas complementary R&D should be taxed. Indeed, in case of substitutable research, foreign R&D is lowered through the provision of a domestic R&D-subsidy. In that case the domestic firm will capture a larger share of the world market since it can produce at lower cost. This in turn will benefit the subsidizing country, thus explaining why the optimal R&D-subsidy is positive for substitutable R&D. For complementary research the reverse holds. In the model employed by Spencer and Brander [1983], firms do not minimize their cost of production. Rather, from a social welfare point of view, each firm invests excessively in R&D.[45] Given then that in the case of complementary R&D foreign R&D is positively affected by a domestic R&D-subsidy it follows that it is optimal to tax this type of research. Third, Spencer and Brander [1983] show that a unilateral provision of an optimal R&D-subsidy induces the subsidized firm to be the leader in the Stackelberg equilibrium of the R&D stage. Finally, if both an R&D-subsidy and an export subsidy can be provided, it is optimal to subsidize exports, in order to capture a larger world market share, and to tax R&D, in order to restore the efficiency of domestic production.

The caveat here is that one country benefits at the expense of the other. As shown by Spencer and Brander [1983], the noncooperative equilibrium in which both countries subsidize their domestic firm entails positive R&D-subsidies. However, in the cooperative equilibrium, in which both countries *jointly* determine the R&D-subsidy, it is optimal to tax R&D. Indeed, if both countries provide an R&D-subsidy, both earn less rent compared to the case in which neither country subsidizes R&D. This being the case, it is doubtful whether the provision of R&D-subsidies to firms operating on international markets is an industrial policy that is to be followed.[46]

As will be shown in Chapter 2 however, the type of product market considered in case of R&D-subsidies matters much more however then conjectured by Spencer and Brander [1983].

[44] Spencer and Brander [1983] consider a balanced budget policy. That is, the cost of subsidization are incurred by the firms (we will discuss this policy in more detail in Chapter 2).

[45] See also Brander and Spencer [1983].

[46] It appears as if the two countries are facing a prisoner's dilemma: subsiding domestic R&D is a dominant strategy if foreign R&D is not subsidized, and the case in which both firms receive an R&D-subsidy is inferior compared to the situation in which both countries abstain from subsidization. However, in order to have a true prisoner's dilemma it also has to be shown that subsidizing domestic R&D is a dominant strategy if foreign R&D is subsidized.

Stochastic R&D

Bagwell and Staiger [1994] introduce uncertainty into the framework developed by Spencer and Brander [1983]. In particular, investments in R&D lead to a reduction in the mean of the investing firm's cost distribution. Bagwell and Staiger [1994] show that all results reported by Spencer and Brander [1983] carry over to this richer framework.[47]

Bagwell and Staiger [1994] consider next international oligopolies. Any government is then confronted with a more complicated policy incentive towards R&D-subsidies, consisting of a strategic and corrective incentive. As before, a larger market share on the world market is obtained through the provision of R&D-subsidies (the strategic incentive), but now any domestic firm's R&D investments also affect any other domestic firm's profits, giving rise the a corrective incentive. If the latter (cross)effect is positive, R&D should be subsidized, whereas an R&D tax should be imposed if the cross-effect is negative. The overall policy incentive towards R&D-subsidies thus depends on the composition of the strategic and corrective effect. Bagwell and Staiger [1983] show that the overall policy incentive is to subsidize (tax) R&D when the cross-effect is positive (negative), in case of international complementary R&D.[48] The optimal R&D policy then consists of a strategic R&D-subsidy and a corrective R&D tax, which resembles the R&D-tax/export-subsidy scheme developed by Spencer and Brander [1983].[49] If R&D is an international substitute, the overall policy incentive remains unclear.

Finally, a more complicated stochastic process is introduced which allows for differences in the 'riskiness' of R&D. In particular, R&D can lower mean cost (the mean effect), but with it the uncertainty as to the emerging distribution of cost (the 'riskiness' of the cost distribution) either increases or decreases (the risk effect). Bagwell and Staiger [1994] show that the optimal R&D policy is highly sensitive to the relative strength of the mean and risk effect. If the mean effect dominates the risk effect, the previous analysis still holds, implying that strategic subsidies and corrective taxes are desired. On the other hand, if the risk effect dominates the mean effect, the optimal policy is the reverse: corrective subsidies and strategic taxes.[50]

---

[47] Observe that deterministic R&D is a special case of stochastic R&D.

[48] R&D is international complementary if $\partial^2 \pi_j^F / \partial x_i^D \partial x_j^F > 0$, where π_j^F are first-stage profits of foreign firm j, and where x_i^D are R&D investments of domestic firm i.

[49] Using examples of a price-setting and quantity-setting duopoly Bagwell and Staiger [1994] illustrate that their findings do not depend on the type of market competition considered.

[50] In a companion paper, Bagwell and Staiger [1992] show that if firms' behaviour is described by a patent-race model, the sensitivity of the optimal R&D policy to the uncertainty of R&D disappears. That is, if firms battle for a monopoly position the optimal R&D policy always consist of strategic subsidies and corrective taxes.

1.7 A policy portfolio: subsidizing cooperative R&D

Industrial policy is sometimes preceded by substantial economic research. In other cases policy makers are ahead of the academic economists. For example, the legislation allowing firms to cooperate in their research was *followed* by an enormous increase in studies devoted to this policy. The same could happen to another widely implemented policy: subsidizing cooperative R&D. To date, this policy has received very little attention, although in modern industrial policy this form of stimulating private R&D is the rule rather than the exception.[51] In this short section the small body of theoretical and empirical literature on the subject is reviewed.

1.7.1 Subsidizing RJV-cartels

To the best of the author's knowledge, the only theoretical research concerned with subsidizing R&D-cooperatives is done by Spence [1984].[52] He briefly considers R&D-subsidies when the whole industry forms an RJV-cartel. Using numerical examples, Spence [1984] concludes that optimal R&D-subsidies substantially increase the incentives of an RJV-cartel to invest in R&D. Moreover, when optimal R&D-subsidies are provided, performance (as defined in Section 1.6.1) improves considerably.

The focus of Spence [1984] however is clearly not on subsidized R&D-cooperatives. As a result, many interesting features of this policy are not discussed. Hence, there is ample scope for additional research in this area.

1.7.2 Empirical evidence

Using a sample of 540 Swedish firms which account for 45 'technologies' Fölster [1995] tests two hypothesis.[53,54] First, do the incentives to cooperate in research (by either trading information, merging, licensing, or forming a joint venture) increase if R&D-subsidies are granted, *provided* that firms participate in some cooperative? Second, are private R&D efforts stimulated by R&D-subsidies?

The answer to the first question is only partly positive. The incentive to form a joint venture is positively influenced by an R&D-subsidy that requires such a cooperative agreement. On the other hand, if firms are allowed to choose between forming a joint venture or trading information, the incentive to choose either form of cooperative agreement is not affected by an R&D-subsidy.

---

[51] Indeed, the European Union spends billions of ECU on subsidizing R&D cooperatives through its framework programs (see European Commission [1995, 1996b]).

[52] Other contributions are Hinloopen [1994 and 1995]. A generalization of these two studies will be presented in Chapter 2.

[53] Technologies are not defined by Fölster [1994]. *Types* of technologies considered are: medical and biotech, communications, energy, environment, information, lasers, new materials, robotics, and transport.

[54] To the best of the author's knowledge, Fölster [1995] is the only empirical study to date considering subsidization of R&D-cooperatives.

As is to be expected, under all types of cooperative agreements considered by Fölster [1995], as well as in the case of noncooperative R&D, providing R&D-subsidies increases private R&D expenditures. On average, each dollar of R&D-subsidy increases private R&D spending with 25 dollarcent.

1.8 Conclusions

There is widespread belief that the market for innovations fails in that market forces do not direct enough resources to innovative activity. This wedge between social and private incentives towards R&D investment is due to the uncertainty surrounding markets for innovations, to the lower private return rate to R&D than the social rate of return, and to pecuniary externalities.

In order to diminish this market failure, authorities have several policy tools at their disposal. Traditional practise is to reward a successful innovator with some market power for a fixed period time. Patents however have several shortcomings. Not only do they imply a non-optimal dissemination of research outcomes, but in practise it appears to be very difficult to dampen technological spillovers. Moreover, large technological spillovers often feature in R&D-intensive industries. Indeed, a firm might well need a sufficiently broad scientific base to assimilate and to use technological spillovers. In sum, the traditional R&D-stimulating policy has come under heavy attack, both from a theoretical and empirical point of view.

A second policy, recently implemented in Europe, the U.S. and Japan, is to allow firms to form R&D-cooperatives. The potential benefits of this policy are threefold. First, technological spillovers are internalized, thus eliminating the free rider problem. Second, R&D often exhibits economies of scale. That is, it might well be that only a consortium of firms has the necessary resources (both financial and physical) to undertake the ever larger, more complex, and more expensive research projects needed these days. Third, also economies of scope characterize the R&D process. Indeed, synergetic effects and risk pooling broadens the research horizon of cooperating firms.

Formal economic models show indeed that cooperative research can lead to an increase in innovative activity. From the ever-growing body of literature on R&D-cooperatives it can be concluded that cooperative R&D exceeds noncooperative R&D whenever pre-cooperative spillovers are large, or when research is complementary (as opposed to substitutable). This would suggest that allowing firms to cooperate in their R&D can be expected to result positive effects when R&D is generic (as opposed to applied).

However, allowing firms to cooperate in R&D has also several drawbacks. First, cooperatives can collectively decide to cut R&D expenses if negative pecuniary externalities prevail, thereby enlarging the gap between actual R&D investments and that desired by a social planner. Second, an agreement to cooperate in R&D could facilitate collusion in other stages of the production process, a harmful reduction in competition which undoubtedly leads to a loss in total net surplus. Finally, R&D-cooperatives can act as a barrier to entry, for instance by setting standards for future applications.

Providing innovating firms with a direct R&D-subsidy is the third widely used policy to stimulate R&D activity. It can be expected that these financial aids indeed stimulate innovation. Moreover, entry barriers will be lower, which could lead to the introduction of more innovative products. On the other hand, the taxes needed to obtain the necessary resources to provide the R&D-subsidy carry a deadweight loss. Also firms may deceive the authorities to obtain the R&D-subsidy, and *a priori* it is not clear that governments will only subsidize successful research projects.

In practise, all three policies are intensively used. Economic analysis considering patents abound. And the last decade has witnessed a substantial increase in the number of studies considering R&D-cooperatives. R&D-subsidies have received very little attention however. Moreover, the literature on subsidized R&D-cooperatives is practically nonexistent. Indeed, much research into this direction is still to be undertaken, given that billions of ECUs are spent on R&D-subsidies, either for competing or cooperating firms.

2 Subsidizing cooperative and noncooperative R&D in a differentiated oligopoly

2.1 Introduction

The literature review presented in Chapter 1 reveals that academic economists have paid little attention to R&D-subsidies. Studies considering both R&D-cooperatives and R&D-subsidies within the same analytical framework are even more scarce. The present chapter is a first step towards filling this gap in economic literature. Setting up an analytical framework with which both R&D-subsidies and different types of R&D-cooperatives can be analyzed not only allows for an evaluation of the relative effectiveness of the two policies, it also paves the way for considering the implementation of both policies *simultaneously*. Given that the latter policy portfolio is more widely implemented, the analysis will be all the more interesting.

The example of d'Aspremont and Jacquemin [1988] is taken as a starting point. The present analysis however allows products to be differentiated so that second-stage Bertrand competition can also be analyzed. Moreover, the industry is assumed to be an oligopoly (recall that d'Aspremont and Jacquemin [1988] only consider two firms). In addition, an active government is introduced which sets an (optimal) R&D-subsidy before the R&D stage. Following Spencer and Brander [1983] it is assumed that in order to finance the R&D-subsidy, firms are taxed for it in the product market. Hence, the set-up allows for a balanced budget policy.

In what follows, first the model is explained in detail. Then equilibria as they emerge under the three scenarios distinguished by d'Aspremont and Jacquemin [1988] are computed for the extended model. Subsequently, the two R&D stimulating policies are considered separately, followed by a comparison of the effectiveness of the two instruments and an analysis of the simultaneous implementation of both policies. Some concluding remarks are stated at the end of the chapter.

2.2 The model

2.2.1 Demand

The demand side of the economy hosts a continuum of consumers of the same type, all represented by a single consumer. This consumer's utility is linear and separable in a *numeraire* good, q_0. The representative consumer maximizes a standard quadratic utility function

$$U(q_0,\dots,q_n) = q_0 + a\sum_{i=1}^{n} q_i - \frac{b}{2}\left[\sum_{i=1}^{n} q_i^2 + \theta\sum_{i=1}^{n} q_i Q_{-i}\right], \tag{2.1}$$

where q_i is the production of commodity i, p_i the concomitant price, $Q_{-i} = \sum_{j=1, \neq i}^{n} q_j$, and a and b

some positive constants. The parameter $\theta \in [0,1]$ captures the extent to which products are differentiated; $\theta = 1$ indicates that all goods are completely homogeneous, while $\theta = 0$ implies that commodities are independent.[1] Observe that it is assumed that all goods are differentiated to the same extent, that is, θ does not vary across commodities.

Given that the representative consumer's budget is equal to $\sum_{i=0}^{n} p_i q_i$, the functional form of utility proposed in (2.1) leads to the following system of inverse demands

$$p_i(q_i, Q_{-i}) = a - b(q_i + \theta Q_{-i}),\tag{2.2}$$

or in direct rather than indirect form

$$q_i(p_i, P_{-i}) = \frac{1}{b(1-\theta)[1+\theta(n-1)]}\{(1-\theta)a + \theta P_{-i} - [1+\theta(n-2)]p_i\},\tag{2.3}$$

where $P_{-i} = \sum_{j=1, \neq i}^{n} p_j$. For (2.3) to be well defined, products cannot be completely homogeneous. Hence, whenever we refer to homogeneous products we implicitly assume products to be differentiated to an infinitesimal extent, that is, $\theta = 1 - \varepsilon$ where $\varepsilon \downarrow 0$.

2.2.2 Supply

The industry consists of a competitive sector, which produces the *numeraire* good, and a monopolistic sector. In the latter each firm produces one variety of the differentiated commodity, q_i. Fixed costs of production are assumed to be the same for all firms and set equal to zero. The marginal cost of production, A, is also constant, but each firm i can lower marginal cost by devoting resources to R&D, x_i. That is, firms which are active in the monopolistic sector can invest in process innovating R&D.

Diminishing returns to scale are inherent in the R&D process (see e.g. Patel and Pavitt [1995]). That is, the relation between a firm's R&D investments and its R&D output is expected to be concave. To capture this property we assume the cost of R&D to be quadratic. Hence, firm i's net benefit of conducting R&D is given by (remember that R&D lowers marginal cost)

$$x_i q_i - \gamma \frac{x_i^2}{2},\tag{2.4}$$

where γ is some positive constant.

It is assumed that the R&D process involves trial and error (see Kamien et al. [1992]). That is, each firm is engaged in pursuing several research paths simultaneously, the difference among firms being the different emphasis each places on each research avenue. In this setting it is unlikely that competitors will conduct identical research. However, during this innovative process firms can learn from each others' experiences, that is, there are technological spillovers. If these are taken into

---

[1] Negative values of θ would indicate that products are complementary. This case, however, is not considered.

account, then firm i's expected net benefit of devoting x_i to R&D is given by

$$(x_i + \beta X_{-i})q_i - \gamma \frac{x_i^2}{2},$$ (2.5)

where $X_{-i} = \sum_{j=1, \neq i}^{n} x_j$ and where $\beta \in [0,1]$ represents the technological spillovers. If $\beta = 0$ there is no leakage of information among competitors, whereas $\beta = 1$ indicates that there is dissemination of knowledge among firms. Observe that (2.5) implies that the amount of technological knowledge that leaks between firms is the same among all firms.

The different types of R&D-cooperatives distinguished by Kamien et al. [1992] (see Section 1.5), are now readily identified. An RJV implies that firms do not cooperate in either stage of the production process. Rather, only β is set equal to one. On the other hand, an R&D-cartel refers to maximization of *joint* profits over R&D investments without, however, affecting the size of β. An RJV-cartel then is an R&D-cartel with $\beta = 1$.

A single firm's (before tax) profit can now be written as

$$\pi_i(p_i, q_i, x_i, X_{-i}) = p_i q_i - (A - x_i - \beta X_{-i})q_i - \gamma \frac{x_i^2}{2}.$$ (2.6)

2.2.3 Government

The authorities have two policy instruments to affect private R&D spending. The first is to allow firms to cooperate in their research by either forming an RJV, an R&D-cartel or an RJV-cartel. These types of cooperatives will be examined in detail in the next sections by solving different types of multi-stage games. Second, following Spencer and Brander [1983], the government can provide a direct subsidy, s, per unit of R&D. In order to analyze this instrument we have to augment the model as presented thus far. It is assumed that each firm receives the same per-unit subsidy. Firm i's before-tax profits with R&D-subsidy are then given by

$$\pi_i^S(p_i, q_i, x_i, X_{-i}) = p_i q_i - (A - x_i - \beta X_{-i})q_i - \gamma \frac{x_i^2}{2} + s x_i.$$ (2.7)

In order to finance the R&D-subsidies, firms are taxed in the product market (see Spencer and Brander [1983]). In equilibrium firm i's after tax profits equal

$$\pi_i^* = (1 - t^*)\pi^{s*} = (1 - t^*)\left[p^* q^* - (A - [1 + \beta(n-1)]x^*)q^* - \gamma \frac{x^{*2}}{2} + s^* x^* \right],$$ (2.8)

where stars denote equilibrium outcomes and where t^* is the equilibrium tax-rate. This tax-rate is

implicitly defined by the net balance of payments

$$NBP = tn\pi_i^{s\bullet} - ns^\bullet x^\bullet. \tag{2.9}$$

Hence, after tax equilibrium profits are given by

$$\pi_i^\bullet = \left(1 - \frac{ns^\bullet x^\bullet + NBP}{n\pi^{s\bullet}}\right)\pi^{s\bullet} = p^\bullet q^\bullet - (A - [1 + \beta(n-1)]x^\bullet)q^\bullet - \gamma\frac{x^{\bullet^2}}{2} - \frac{NBP}{n}. \tag{2.10}$$

Throughout this chapter we confine ourselves to a balanced budget policy in order to 'honestly' assess the effects of R&D-subsidies. In that case no money 'just' enters or leaves the industry. Note that with a zero *NBP* firms' after tax equilibrium profits *may* look like equilibrium profits without any government intervention (compare (2.6) with (2.10)). However, by providing R&D-subsidies and taxing production accordingly, the government changes the cost structure of the R&D stage, and thus changes the set of actions which are compatible with the two-stage Nash-Cournot equilibrium. Firms cannot alter the cost structure themselves since, by definition of a Nash-Cournot equilibrium, it is not in their interest to shift financial resources from the output stage to the R&D stage (see also Spencer and Brander [1983]).

Both policy instruments are assessed according to their effect on social welfare. In equilibrium, consumers' surplus in the market considered here is given by[2]

$$CS = U(q^\bullet) - np^\bullet q^\bullet = n\left(a - \frac{b}{2}[1 + \theta(n-1)]q^\bullet - p^\bullet\right)q^\bullet, \tag{2.11}$$

and producers' surplus equals

$$PS = n(1 - t^\bullet)\pi^{s\bullet}. \tag{2.12}$$

Hence, social welfare, defined as the sum of producers' and consumers' surplus *and* the NBP, is[3]

$$W = n\left\{\left(a - \frac{b}{2}[1 + \theta(n-1)]q^\bullet\right)q^\bullet - (A - [1 + \beta(n-1)]x^\bullet)q^\bullet - \gamma\frac{x^{\bullet^2}}{2}\right\}. \tag{2.13}$$

It is this expression that the government seeks to maximize when providing R&D-subsidies, supposing that firms' output and R&D investments are the equilibrium outcomes of the appropriate market game.

---

[2] See Spence [1976]. Measuring consumers' surplus when products are differentiated can be difficult (see the comment of Wildman [1984] on Scherer [1979]).

[3] Strictly speaking we do not have to include the NBP since it is assumed to be zero.

2.3 Market equilibria[4]

Having described the model we employ to assess the two R&D stimulating policies we now proceed with computing the market equilibria as they arise under the three regimes considered by d'Aspremont and Jacquemin [1988].

2.3.1 No cooperation in either R&D or production

In the second stage of the game firms compete over either price or quantities. In the latter case each firm i maximizes (2.7) over q_i, which results in the equilibrium quantity conditional on R&D efforts[5]

$$\hat{q}_i^{\,C}(x_i,X_{-i}) = \frac{1}{b(2-\theta)[2+\theta(n-1)]}$$
$$\times \left\{(a-A)(2-\theta)+\{2+\theta(n-2)-\theta\beta(n-1)\}x_i+(2\beta-\theta)X_{-i}\right\}. \tag{2.14}$$

On the other hand, in case of second-stage Bertrand competition, each firm i maximizes (2.7) with respect to p_i. Equilibrium prices thus derived equal

$$\hat{p}_i^{\,B}(x_i,X_{-i}) = \frac{1}{[2+\theta(n-3)][2+\theta(2n-3)]}$$
$$\times \left\{\alpha_1^B-[1+\theta(n-2)][\{2+\theta(n-2)+\theta\beta(n-1)\}x_i+\{\theta+2\beta[1+\theta(n-2)]\}X_{-i}]\right\}, \tag{2.15}$$

with $\alpha_1^B = [2+\theta(2n-3)]\{A[1+\theta(n-2)]+a(1-\theta)\}$.

In the preceding stage, when firms determine their R&D investments, profits in case of second-stage Cournot or Bertrand competition can be written respectively as

$$\hat{\pi}_i^C(x_i,X_{-i};s) = \frac{[\alpha_2^C]^2}{b(2-\theta)^2[2+\theta(n-1)]^2} -\gamma\frac{x_i^2}{2}+sx_i, \tag{2.16a}$$

with $\alpha_2^C = (a-A)(2-\theta)+\{2+\theta(n-2)-\theta\beta(n-1)\}x_i+(2\beta-\theta)X_{-i}$, and

$$\hat{\pi}_i^B(x_i,X_{-i};s) = \frac{\{\hat{p}_i^{\,B}(x_i,X_{-i})-(A-x_i-\beta X_{-i})\}\alpha_2^B}{b(1-\theta)[1+\theta(n-1)]} -\gamma\frac{x_i^2}{2}+sx_i, \tag{2.16b}$$

with $\alpha_2^B = (1-\theta)a-[1+\theta(n-2)]\hat{p}_i^{\,B}(x_i,X_{-i})+\theta\hat{P}_{-i}^{\,B}(x_i,X_{-i})$. Maximizing (2.16) with respect to the

---

[4] All analytical expressions presented in this chapter are numerically checked. GAUSS-programs with which these checks are performed, are available upon request.

[5] A hat refers to a conditional equilibrium. Second stage Cournot competition is denoted by C, whereas second stage Bertrand competition is indicated by B.

R&D efforts results in the following R&D reaction functions[6]

$$x_i^C(x_j^C) = \frac{\alpha_3^C + 2(n-1)(2\beta-\theta)\{2+\theta(n-2)-\theta\beta(n-1)\}x_j^C}{b\gamma(2-\theta)^2[2+\theta(n-1)]^2 - 2\{2+\theta(n-2)-\theta\beta(n-1)\}^2}. \tag{2.17a}$$

where $\alpha_3^C = 2(a-A)(2-\theta)\{2+\theta[(n-2)-\beta(n-1)]\} + b(2-\theta)^2[2+\theta(n-1)]^2 s$, and

$$x_i^B(x_j^B) = \frac{2\Delta[1+\theta(n-2)]\{\alpha_3^B + \{(2\beta-\theta)[1+\theta(n-2)]-\theta^2\beta(n-1)\}(n-1)x_j^B}{b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2[2+\theta(2n-3)]^2 - 2\Delta^2[1+\theta(n-2)]}. \tag{2.17b}$$

with $\alpha_3^B = (a-A)(1-\theta)[2+\theta(2n-3)] + b(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2[2+\theta(2n-3)]^2 s$, and where $\Delta = 2+3\theta(n-2)+\theta^2[(n-1)(n-2)-(2n-3)]-\theta\beta(n-1)[1+\theta(n-2)]$.[7] From these we derive the following stability conditions

$$b\gamma(2-\theta)[2+\theta(n-1)]^2 > 2\{2+\theta\{(n-2)-\beta(n-1)\}[1+\beta(n-1)]\}, \tag{2.18a}$$

and

$$b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 > 2\Delta[1+\theta(n-2)][1+\beta(n-1)] \tag{2.18b}$$

Equating the R&D reaction functions leads to the conditional equilibrium R&D efforts

$$\hat{x}_{IC}(s) = \frac{2(a-A)\{2+\theta[(n-2)-\beta(n-1)]\} + b(2-\theta)[2+\theta(n-1)]^2 s}{b\gamma(2-\theta)[2+\theta(n-1)]^2 - 2\{2+\theta[(n-2)-\beta(n-1)]\}[1+\beta(n-1)]}. \tag{2.19a}$$

and

$$\hat{x}_{IB}(s) = \frac{2\Delta(a-A)[1+\theta(n-2)] + b[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 s}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 - 2\Delta[1+\theta(n-2)][1+\beta(n-1)]}. \tag{2.19b}$$

where *I* refers to the fully noncooperative nature of the game considered.

As is to be expected, given the second-order conditions associated with finding the optimal R&D investments, under both second-stage Cournot and Bertrand competition, competitive R&D efforts are increasing in the R&D-subsidy.

---

[6] The second-order condition for a global maximum under second-stage Cournot competition is $b\gamma(2-\theta)^2[2+\theta(n-1)]^2 > 2\{2+\theta[\theta(n-2)-\beta(n-1)]\}^2$, whereas under Bertrand competition in the product market it is $b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2[2+\theta(2n-3)]^2 > 2\Delta^2[1+\theta(n-2)]$.

[7] Observe that $\Delta>0$, $\forall n, \beta, \theta$.

Social welfare under the two types of product market competition considered is given by

$$\frac{\hat{W}^C(s)}{n} = \frac{3+\theta(n-1)}{2b[2+\theta(n-1)]^2}\{(a-A)+[1+\beta(n-1)]\hat{x}^C(s)\}^2-\gamma\frac{\hat{x}^C(s)^2}{2},$$ (2.20a)

and

$$\frac{\hat{W}^B(s)}{n} = \frac{[1+\theta(n-2)][3+\theta(n-4)]}{2b[1+\theta(n-1)][2+\theta(n-3)]^2}\{(a-A)+[1+\beta(n-1)]\hat{x}^B(s)\}^2-\gamma\frac{\hat{x}^B(s)^2}{2},$$ (2.20b)

respectively. Maximizing (2.20) with respect to s results in the optimal R&D-subsidies for the fully competitive games[8]

$$s_{IC}^* = \frac{\gamma(a-A)\{(2-\theta)[3+\theta(n-1)][1+\beta(n-1)]-2\{2+\theta[(n-2)-\beta(n-1)]\}\}}{(2-\theta)\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}},$$ (2.21a)

and

$$s_{IB}^* = \frac{\gamma(a-A)[1+\theta(n-2)]}{2+\theta(2n-3)}$$

$$\times \frac{[2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-2\Delta}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}.$$ (2.21b)

The analytical equilibrium expressions for all relevant variables are summarized in Tables 2A.1 and 2A.3 of Appendix 2A. Given the computed equilibrium, the equilibrium tax rate is defined by (2.9). Under second-stage Cournot and Bertrand competition these are

$$t_{IC}^* = 2[1+\beta(n-1)][3+\theta(n-1)]$$

$$\times \frac{(2-\theta)[3+\theta(n-1)][1+\beta(n-1)]-2[2+\theta\{(n-2)-\beta(n-1)\}]}{2b\gamma(2-\theta)[2+\theta(n-1)]^2-[1+\beta(n-1)][3+\theta(n-1)]\alpha_4^C},$$ (2.22a)

with $\alpha_4^C=2+\theta(2n-3)+\theta^2(n-1)-\beta(n-1)[6+\theta(2n-1)-\theta^2(n-1)]$, and

$$t_{IB} = 2[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]$$

$$\times \frac{[2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-2\Delta}{2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2+\alpha_{41}^B\alpha_{42}^B},$$ (2.22b)

with $\alpha_{41}^B=[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]$, and $\alpha_{42}^B=[2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-4\Delta$, respectively. In order not to impose a tax on production that exceeds gross profits, (2.22) cannot

---

[8] Under Cournot competition the second-order condition is $b\gamma[2+\theta(n-1)]^2>[3+\theta(n-1)][1+\beta(n-1)]^2$, while $b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2>[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2$ is the sufficient condition under Bertrand competition.

exceed 1. This condition translates into

$$2b\gamma[2+\theta(n-1)]^2 > [3+\theta(n-1)]^2[1+\beta(n-1)]^2,$$ (2.23a)

and

$$2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2 > [1+\theta(n-2)][3+\theta(n-4)]^2[1+\beta(n-1)]^2,$$ (2.23b)

respectively.

2.3.2 Cooperation in R&D, competition in production

If firms are allowed to cooperate in the innovation stage, but still compete when producing with the new technology, the second stage of the game, conditional on the R&D efforts, is not different from that under the full competition regime. Hence, if firms compete over quantities, their optimal response is still given by (2.16), whereas (2.17) is the optimal price to quote if firms face second-stage price competition.

In the present scenario however, firms maximize *joint* profits in first stage. These are

$$\hat{\Pi}^C(x;s) = \frac{1}{b(2-\theta)^2[2+\theta(n-1)]^2}$$

(2.24a)

$$\times \sum_{i=1}^{n}\left\{\{(a-A)(2-\theta)+\{2+\theta[(n-2)-\beta(n-1)]\}x_i+(2\beta-\theta)X_{-i}\}^2-\frac{\gamma x_i^2}{2}+sx_i\right\},$$

or

$$\hat{\Pi}^B(x;s) = \frac{1}{b(1-\theta)[1+\theta(n-1)]}$$

$$\times \sum_{i=1}^{n}\left\{\{\hat{p}_i^B(x_i,X_{-i})-(A-x_i-\beta X_{-i})\}\right.$$ (2.24b)

$$\times \left.\{(1-\theta)a-[1+\theta(n-2)]\hat{p}_i^B(x_i,X_{-i})+\theta\hat{P}_{-i}^B(x_i,X_{-i})\}-\gamma\frac{x_i^2}{2}+sx_i\}\right\},$$

under second-stage Cournot and Bertrand competition respectively. Optimal R&D efforts, which follow from maximizing (2.24) over x_i, are given by[9]

---

[9] The second-order condition is $b\gamma[2+\theta(n-1)]^2>2[1+\beta(n-1)]^2$ under Cournot competition, while under Bertrand competition it reads $b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2>2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2$.

$$\hat{x}_{IIC}(s) = \frac{2(a-A)[1+\beta(n-1)]+b[2+\theta(n-1)]^2 s}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2}, \quad (2.25a)$$

and

$$\hat{x}_{IIB}(s) = \frac{2(a-A)(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]+b[1+\theta(n-1)][2+\theta(n-3)]^2 s}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}, \quad (2.25b)$$

where II indicates that the game is partially cooperative.

As in the case of noncooperative R&D, private R&D efforts are positively influenced by a positive R&D-subsidy, given the second-order conditions associated with the second stage of the game.

Observe that since the second stage of the game is the same as under the fully competitive regime, social welfare is still given by (2.20), with, however the respective conditional equilibrium R&D efforts given by (2.25). Maximizing then these expressions for social welfare with respect to s leads to the optimal R&D-subsidies for the partially cooperative games. Under second-stage Cournot and Bertrand competition these are[10]

$$s_{IIB}^{*} = \frac{\gamma(a-A)[1+\theta(n-1)][1+\beta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}, \quad (2.26a)$$

and

$$s_{IIB}^{*} = \frac{\gamma(a-A)[1+\theta(n-2)]^2[1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}, \quad (2.26b)$$

respectively.

In Tables 2A.2 and 2A.4 of Appendix 2A all analytical equilibrium expressions of the game considered in this section are presented. The concomitant equilibrium tax rates under second-stage Cournot and Bertrand competition are respectively given by

$$t_{IIC}^{*} = \frac{2[1+\theta(n-1)][3+\theta(n-1)][1+\beta(n-1)]^2}{2b\gamma[2+\theta(n-1)]^2-[1-\theta(n-1)][3+\theta(n-1)][1+\beta(n-1)]^2}, \quad (2.27a)$$

and

$$t_{IIB}^{*} = \frac{2[3+\theta(n-4)][1+\theta(n-2)]^2[1+\beta(n-1)]^2}{2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2+(\theta n-1)[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}. \quad (2.27b)$$

Again, firms' after-tax profits cannot go negative. It turns out that the concomitant requirement on the equilibrium tax rates, that these are below 1, leads to the same conditions as those stated in (2.23).

---

[10] Under second-stage Cournot competition $b\gamma[2+\theta(n-1)]^2>[3+\theta(n-1)][1+\beta(n-1)]^2$ is the second-order condition, while $b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2>[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2$ is required under second-stage Bertrand competition.

2.3.3 Cooperation in both R&D and production

If all firms cooperate in determining their production as well as in setting their R&D efforts, they collectively act as a monopolist. In that case there is no difference in the type of market competition considered since there is no competitive production stage. Hence, if production is jointly set, then the conditional equilibrium price follows immediately from (2.2). Likewise, conditional equilibrium quantities are defined by (2.3) if price is the second stage decision variable. In what follows we treat the cooperative as determining the optimal quantity in the second stage.

Joint second-stage profits are

$$\Pi_{III}(q_i, Q_{-i}; x_i, X_{-i}, s) = \sum_{i=1}^{n} \left\{ (a - bq_i - \theta b Q_{-i}) q_i - (A - x_i - \beta X_{-i}) q_i - \gamma \frac{x_i^2}{2} + s x_i \right\}. \tag{2.28}$$

Equilibrium quantities, conditional on the monopoly R&D effort, are obtained by maximizing (2.28) over q_i. This results in

$$\hat{q}_{III}(x) = \frac{(a - A) + [1 + \beta(n-1)]x}{2b[1 + \theta(n-1)]}. \tag{2.29}$$

Joint first-stage profits are then given by

$$\hat{\Pi}_{III}(x; s) = n \left\{ \frac{1}{4b[1 + \theta(n-1)]} \{(a - A) + [1 + \beta(n-1)]x\}^2 - \frac{\gamma x^2}{2} + s x \right\}. \tag{2.30}$$

In the first stage the monopoly maximizes (2.30) to obtain its optimal R&D effort. Conditional upon the R&D-subsidy, this effort is given by[11]

$$\hat{x}_{III}(s) = \frac{(a - A)[1 + \beta(n-1)] + 2b[1 + \theta(n-1)]s}{2b\gamma[1 + \theta(n-1)] - [1 + \beta(n-1)]^2}. \tag{2.31}$$

Also under the fully cooperative regime, R&D efforts are increasing in the R&D-subsidy. It remains to be analyzed however if R&D-subsidies will indeed be positive under the three regimes considered here, an issue that will be discussed extensively in Section 2.5.

---

[11] The second-order condition reads as $2b\gamma[1 + \theta(n-1)] > [1 + \beta(n-1)]^2$.

Social welfare in case of a monopoly can be written as

$$\frac{\hat{W}_{III}(s)}{n} = \frac{3}{8b[1+\theta(n-1)]}\{(a-A)+[1+\beta(n-1)]\hat{x}_{III}(s)\}^2 - \gamma\frac{\hat{x}_{III}(s)^2}{2}.$$ (2.32)

The optimal R&D-subsidy maximizes this expression and is given by[12]

$$s_{III}^* = \frac{\gamma(a-A)[1+\beta(n-1)]}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}.$$ (2.33)

Analytical equilibrium expressions for all relevant variables in case of a monopoly are presented in Table 2A.5 of Appendix 2A. The equilibrium tax rate under monopoly is

$$t_{III}^* = \frac{6[1+\beta(n-1)]^2}{8b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}.$$ (2.34)

Imposing that this tax rate does not exceed 1 leads to

$$8b\gamma[1+\theta(n-1)] > 9[1+\beta(n-1)]^2.$$ (2.35)

2.4 A case for cooperation in R&D?

Having computed the equilibria as they arise under the three regimes considered the effects of firms cooperating in one or two stages of the production process on market performance can be evaluated (R&D-subsidies will be considered in the next section). In order to do so, first the influences on R&D efforts of the different types of cooperatives are examined. Then the social incentives towards cooperation in either R&D or both R&D and production are contrasted with private incentives. The conclusion of this section will be that there is a mild case to be made for sustaining R&D-cooperatives. Indeed, R&D-cooperatives can result in increased R&D activity. However, they can also lead to a reduction in private spending on R&D. Moreover, the threat of extending the R&D-cooperative agreement to the production stage is clearly present.

2.4.1 Cooperative research and development

The complete ranking of private R&D efforts is described in Lemma 2.1 (the proof of which is given in Appendix 2B.1; there it is also shown that the regions defined in the lemma in terms of the technological spillovers are well defined).

---

[12] The second-order condition states that $4b\gamma[1+\theta(n-1)]>3[1+\beta(n-1)]^2$.

Lemma 2.1

$\forall\, n, \beta, \theta$, the following holds

$$\hat{x}_{IC}(0) > \hat{x}_{III}(0) > \hat{x}_{IIC}(0) \quad \Leftrightarrow \quad \beta \in [0, C_1],$$

$$\hat{x}_{III}(0) > \hat{x}_{IC}(0) > \hat{x}_{IIC}(0) \quad \Leftrightarrow \quad \beta \in (C_1, C_2),$$

$$\hat{x}_{III}(0) > \hat{x}_{IIC}(0) > \hat{x}_{IC}(0) \quad \Leftrightarrow \quad \beta \in (C_2, 1],$$

with $C_1 = \theta[4 + 2\theta(n-1) + \theta^2(n-1)]/\{8 + 8\theta(n-1) + 2\theta^2(n-1)^2 - \theta^3(n-1)^2\}$ and $C_2 = \theta/2$, and

$$\hat{x}_{IB}(0) > \hat{x}_{III}(0) > \hat{x}_{IIB}(0) \quad \Leftrightarrow \quad \beta \in [0, B_1),$$

$$\hat{x}_{III}(0) > \hat{x}_{IB}(0) > \hat{x}_{IIB}(0) \quad \Leftrightarrow \quad \beta \in (B_1, B_2),$$

$$\hat{x}_{III}(0) > \hat{x}_{IIB}(0) > \hat{x}_{IB}(0) \quad \Leftrightarrow \quad \beta \in (B_2, 1],$$

with $B_1 = \theta[4 + 2\theta(3n-7) + \theta^2(2n^2 - 11n + 13)]/\{8 + 16\theta(n-2) + 2\theta^2[5n^2 - 20n + 19] + \theta^3(n-1)(2n^2 - 9n + 11)\}$, and $B_2 = \theta[1 + \theta(n-2)]/\{2 + 2\theta(n-2) - \theta^2(n-1)\}$.

The lemma is illustrated in Figures 2.1a and 2.1b. The regions in the figures are labelled according to the value of the spillover parameter. In particular, β_1^C refers to $\beta \in [0, C_1)$, β_2^C implies $\beta \in (C_1, C_2)$, and β_3^C points to $\beta \in (C_2, 1]$. A similar notation applies for Bertrand competition.

Recall from Section 1.5 that the effect of allowing firms to cooperate in R&D (but not in production) on private R&D spending is determined by the interaction between the combined-profits externality and the competitive advantage externality (see also Figure 1.2). It was explained that under second-stage Cournot competition the critical size of technological spillover (that is, the minimum size of pre-cooperative technological spillover required for the combined-profits externality to off set the competitive-advantage externality) is a decreasing function of the extent to which products are differentiated. According to Lemma 2.1 the same is true under product market price competition. A fundamental difference, however, is that under second-stage Cournot competition the critical spillover is independent of the number of firms active in the industry, whereas under second-stage Bertrand competition it is increasing in the number of firms cooperating in their R&D (that is, β^* in Figure 1.2 moves to the right when n increases).

The explanation of this difference lies in the influence of the number of firms in the industry on the two externalities at work. If this number increases, the line in Figure 1.2 representing the competitive-advantage externality rotates clockwise. Indeed, the more firms are active in the industry, the more diluting the competitive-advantage externality on individual incentives to invest in R&D. On the other hand, the line illustrating the combined-profits externality rotates counterclockwise if more firms make up the industry. That is, the more firms are present, the more likely the combined-profits
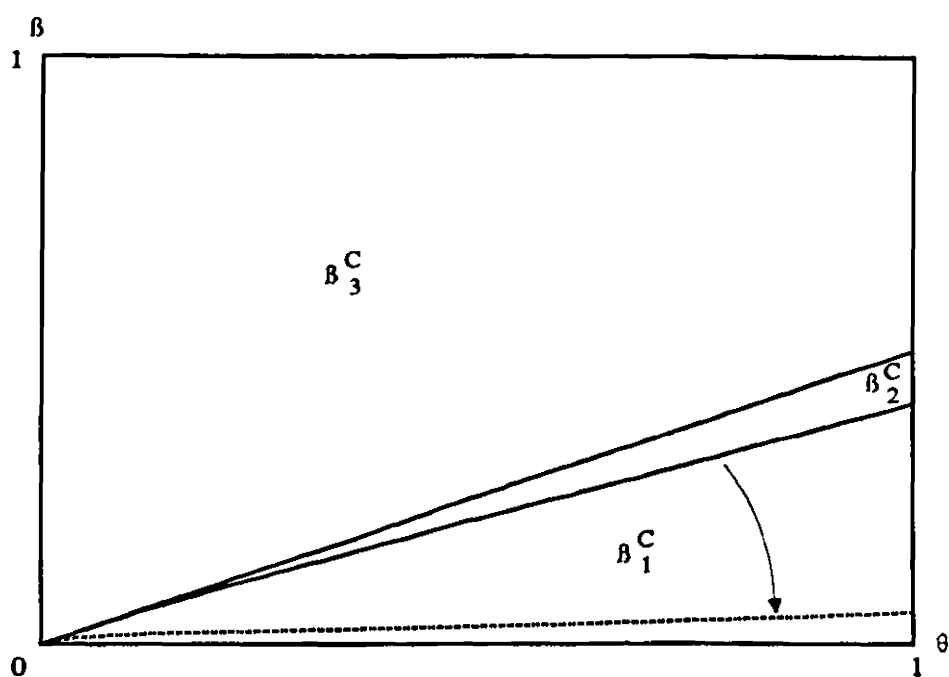
Figure 2.1a   Regions dividing R&D investments for different regimes; Cournot competition
                Effect of increasing *n* from *n*=2 to *n*=50
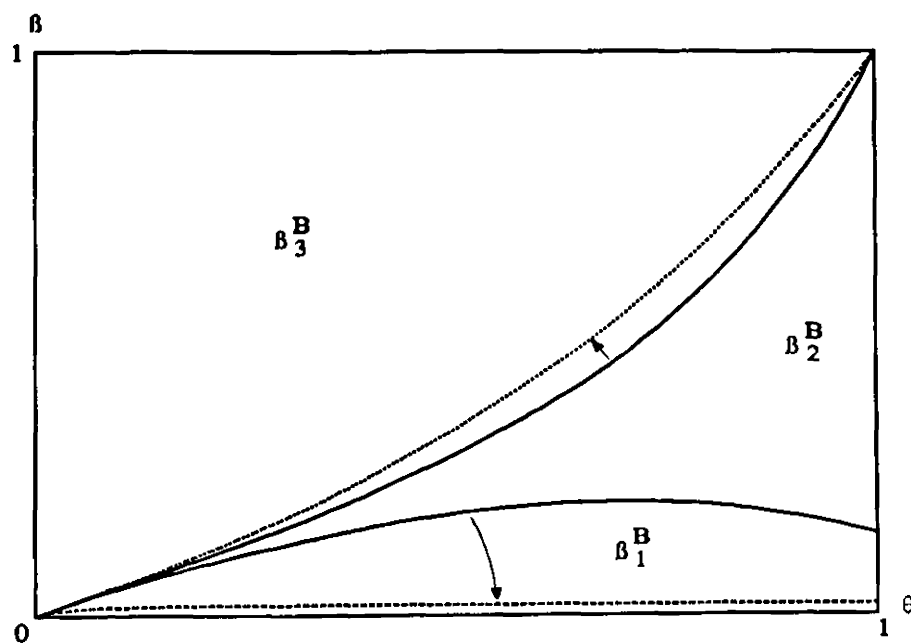
Figure 2.1b   Regions dividing R&D investment for different regimes; Bertrand competition
                Effect of increasing *n* from *n*=2 to *n*=50

externality is positive, since more firms benefit from a single firm's R&D activities. The relative magnitude of this rotation depends on the type of product market competition. According to Lemma 2.1, it exactly offsets the change in the competitive-advantage externality if firms compete over quantities. In case of second-stage Bertrand competition however, the competitive-advantage externality reacts more to an increase in n than the combined-profits externality. The reason for this stronger reaction is that in this case more of the surplus generated through R&D activities is directed towards consumers. Accordingly, under second-stage competition over price, any firms' R&D efforts are more likely to reduce total industry profits, and the effect of an increased number of firms in the industry is an increase in the critical spillover parameter.

A second mechanism revealed by Lemma 2.1 is that increasing the number of firms enlarges the parameter space in which R&D efforts of a monopolist are the highest. Indeed, for n going to infinity a monopolist always invests most in R&D.[13] As observed by d'Aspremont and Jacquemin [1988], this is due to firms cooperating in the product market being more able to capture the surplus generated by their research, compared to either the partial or full competitive regime. The larger the number of firms in the industry, the lower profits are because of more intensive competition. Hence, the more significant the surplus-capturing effect, due to product market cooperation, will be.

Lemma 2.1. can also be interpreted if complementary and substitutable research is distinguished. As shown by Bulow et al. [1985], R&D is complementary *iff.* $\partial^2 \hat{\pi}_I(0)/\partial x_i \partial x_j$ is positive, and substitutable *iff.* $\partial^2 \hat{\pi}_I(0)/\partial x_i \partial x_j$ is negative. Observe that in the model employed here, the appropriate first-order conditions read as

$$\frac{\partial \hat{\pi}_{IC}(0)}{\partial x_i} = \alpha_5^C \{(a-A)(2-\theta)+[(2-\theta)+\theta(1-\beta)(n-1)]x_i+(2\beta-\theta)X_{-i}\}$$

$$-\gamma x_i+s,$$

(2.36a)

and

$$\frac{\partial \hat{\pi}_{IB}(0)}{\partial x_i} = \alpha_5^B \{(a-A)(1-\theta)[2+\theta(2n-3)]+\Delta x_i+[(2\beta-\theta)[1+\theta(n-2)]X_{-i}\}$$

$$-\gamma x_i+s,$$

(2.36b)

where α_5^C and α_5^B are some positive constants. From (2.36) it is straightforward to derive that

$$\frac{\partial^2 \hat{\pi}_{IC}(0)}{\partial x_i \partial x_j} < 0 \quad \Leftrightarrow \quad \beta < C_2,$$

(2.37a)

and

---

[13] Using l'Hospital's rule it is straightforward to show that $\lim_{n \to \infty} C_1 = \lim_{n \to \infty} \theta(2+\theta)/\{8+2\theta(2-\theta)(n-1)\} = 0$, and $\lim_{n \to \infty} B_1 = \lim_{n \to \infty} 2\theta/\{10+\theta(6n-11)\} = 0$.

$$\frac{\partial^2 \hat{\pi}_{IB}(0)}{\partial x_i \partial x_j} < 0 \quad \Leftrightarrow \quad \beta < B_2, \tag{2.37b}$$

where C_2 and B_2 are as defined in Lemma 2.1. It thus follows that also in the model used here, competitive research efforts exceed those of an R&D-cartel whenever R&D is substitutable. The reverse holds for complementary research (see also Section 1.5).

The following lemma, proved in Appendix 2B.2, completes the ranking of R&D efforts for the different types if R&D-cooperatives considered.

Lemma 2.2

part (i)

$\forall n, \beta, \theta$, the following holds

$$\hat{x}_{IIC}(0)|_{\beta=1} > \hat{x}_{IC}(0), \hat{x}_{IIC}(0),$$

$$\hat{x}_{IIB}(0)|_{\beta=1} > \hat{x}_{IIB}(0).$$

If $\theta < 2/3$, $\forall n, \beta$ the following holds

$$\hat{x}_{IIB}(0)|_{\beta=1} > \hat{x}_{IB}(0).$$

part (ii)

$\forall n, \beta, \theta$, the following holds

$$\hat{x}_{III}(0)|_{\beta=1} > \hat{x}_{IIC}(0), \hat{x}_{IIB}(0), \hat{x}_{III}(0).$$

As discussed in Section 1.5, Kamien et al. [1992] show that under second-stage Cournot competition an RJV-cartel results in the highest R&D investments of all R&D-cooperatives considered. In addition, Lemma 2.2 shows that the incentives to invest in R&D are even more enhanced through the rent-capturing effect of cooperation in the product market.[14]

Under second-stage Bertrand competition, a sufficient condition for an RJV-cartel to devote more resources to R&D compared to the competitive alternative is that products are differentiated at least to a modest extent. In that case, the increased force of the combined-profits externality due to full information sharing outweighs the increased diluting effect on R&D investments of the competitive-advantage externality. Indeed, as discussed in Section 1.5, the more products are

---

[14] This result is also mentioned by Kamien et. al [1992].

differentiated, the more important the combined-profits externality will be, and the less important the competitive-advantage externality. On the other hand, there are situations in which competitive R&D efforts exceed those of an RJV-cartel. This could only happen if products are quasi-homogenous and pre-cooperative spillovers small. Under these circumstances any firm's R&D efforts considerably reduce all other firms' profits because of the similarity of products and the intense product market competition. Hence, the combined-profits externality is more likely to be negative or only mildly positive. In addition, full exchange of information has a relatively large effect on each firm's incentive to invest in R&D since pre-cooperative spillovers are small. Together this means that competitive R&D efforts could exceed those of an RJV-cartel. Observe however, that $\theta > 2/3$ is only a *sufficient* condition. That is, also under second-stage Bertrand competition, in most cases an RJV-cartel will yield the highest investment in R&D.

Whether or not an RJV-cartel with cooperation in the product market can be an endogenous market outcome however is a question to be considered next.

2.4.2 Private incentives towards cooperation in R&D

The phrase 'allowing firms to cooperate in R&D' does not make sense if firms do not want to set R&D investments cooperatively. Indeed, any R&D-cooperative can settle for the noncooperative solution if this is more profitable. The private incentives towards collusive agreements are summarized in the following lemma (proved in Appendix 2B.3).

Lemma 2.3

$\forall\, n, \beta, \theta$, the following holds

$$\hat{\pi}_{III}(0) > \hat{\pi}_{IIC}(0) > \hat{\pi}_{IC}(0) \quad \Leftrightarrow \quad \beta \in [0, C_2) \cup (C_2, 1],$$

$$\hat{\pi}_{III}(0) > \hat{\pi}_{IIC}(0) = \hat{\pi}_{IC}(0) \quad \Leftrightarrow \quad \beta = C_2,$$

and

$$\hat{\pi}_{III}(0) > \hat{\pi}_{IIB}(0) > \hat{\pi}_{IB}(0) \quad \Leftrightarrow \quad \beta \in [0, B_2) \cup (B_2, 0],$$

$$\hat{\pi}_{III}(0) > \hat{\pi}_{IIB}(0) = \hat{\pi}_{IB}(0) \quad \Leftrightarrow \quad \beta = B_2,$$

where C_2 and B_2 are defined in Lemma 2.1.

Under both second-stage Cournot and Bertrand competition, and irrespective of the size of technological spillovers and the extent to which products are differentiated, firms always want to collude in as many stages of the production process as allowed. Hence, the endogenous market outcome will be a monopoly.

An R&D-cooperative agreement also rules the extent to which firms share innovative information. Part of the concomitant incentives are summarized in Lemma 2.4, the proof of which can be found in Appendix 2B.4.

Lemma 2.4

$\forall\, n, \beta, \theta$, the following holds

$$\frac{\partial \hat{\pi}_{IIC}(0)}{\partial \beta}, \frac{\partial \hat{\pi}_{IIB}(0)}{\partial \beta}, \frac{\partial \hat{\pi}_{III}(0)}{\partial \beta} > 0.$$

If firms form an R&D-cartel, they always want to exchange innovative information as much as possible. Under second-stage Cournot competition, Katz [1986] reports similar findings for an R&D-cartel without product market collusion. According to Lemma 2.4 this information-sharing incentive is independent of the type of product market competition, and whether or not firms cooperate in the second stage of the production process. That is, if firms cooperate in R&D, or both in R&D and production, there is an incentive to fully share innovative information within the R&D-cooperative.

Private incentives to form an RJV are less clear cut. Simulation results presented in Table 2D.1 of Appendix 2D.2 show that competing firms' profits are increasing in the spillover rate if both the number of firms in the industry and pre-cooperative spillovers are not too high.[15] In that case, each firm benefits more from all information it receives from its competitors vis-à-vis the information it disseminates in the industry. On the other hand, if the number of firms benefiting from any firms' innovative information is large, and if there are substantial pre-cooperative spillovers, full information sharing will lower individual firms' profits.

Together Lemmata 2.2, 2.3 and 2.4 imply that market forces generate the highest amount of R&D investments possible. That is, if firms are allowed to cooperate in as many stages as desired, an industry-wide monopoly will arise in which innovative information is fully shared, a market structure which yields the highest R&D investments. Whether or not this market outcome is desired from a social welfare point of view is considered next.

2.4.3 Social incentives towards cooperation in R&D

The consequences for net total surplus of allowing firms to engage in either of the R&D-cooperatives considered, are revealed through a number of lemmata and some numerical simulations. The first lemma summarizes the effect of an R&D-cartel.

---

[15] Whenever possible, analytical expressions are presented. However, sometimes these are too complicated to be interpreted, or to be derived in the first place. In these cases, numerical approximations are used. In Appendix 2D.1 the used procedures are described.

Lemma 2.5

$\forall n, \beta, \theta$, the following holds

$$\hat{W}_{IB}(0) > \hat{W}_{IIB}(0) \quad \Leftrightarrow \quad \beta \in [0, B_2),$$

$$\hat{W}_{IB}(0) = \hat{W}_{IIB}(0) \quad \Leftrightarrow \quad \beta = B_2,$$

$$\hat{W}_{IB}(0) < \hat{W}_{IIB}(0) \quad \Leftrightarrow \quad \beta \in (B_2, 1],$$

where B_2 is defined in Lemma 2.1. If $n=2$, or if $n \geq 3 \wedge \beta > 1/12$, $\forall \theta$ the following holds

$$\hat{W}_{IC}(0) > \hat{W}_{IIC}(0) \quad \Leftrightarrow \quad \beta \in [0, C_2],$$

$$\hat{W}_{IC}(0) = \hat{W}_{IIC}(0) \quad \Leftrightarrow \quad \beta = C_2,$$

$$\hat{W}_{IC}(0) < \hat{W}_{IIC}(0) \quad \Leftrightarrow \quad \beta \in (C_2, 1],$$

where C_2 is defined in Lemma 2.1.

According to Lemma 2.5, allowing firms to form an R&D-cartel is only desired if this leads to an increase in total R&D efforts. From the discussion in Section 2.4.1 it is clear that this occurs if and only if research is complementary. That is, R&D-cartels enhance net total surplus if participants of the R&D-cooperative are engaged in mutually complementary research. On the other hand, cooperative agreements between firms engaged in substitutable research will yield a reduction in total R&D spending, and with it a reduction in total net surplus.

To put it differently, if the competitive-advantage externality is outweighed by the combined-profits externality, firms collectively decide to spend more on R&D, which leads to an increase in net total surplus. Recall that the combined-profits externality is more likely to overrule the competitive-advantage externality if the extent to which products are differentiated increases (see Section 1.5). That is, if θ falls, both C_2 and B_2 fall, implying that the formation of an R&D-cartel leads to an increase in resources devoted to R&D. Hence, the more distinct products are, the more likely the formation of an R&D-cartel will lead to an increase in net total surplus.

Observe that under second-stage Cournot competition the pre-cooperative spillover rate has to be above a threshold of $1/12$ for the ranking to go through (given that the number of firms in the industry is at least 3). For values of C_2 below $1/12$ it thus always pays society to allow for R&D-cartels if the pre-cooperative spillovers exceed the threshold. This condition then translates to $\theta < 1/6$. That is, if pre-cooperative spillovers are modest and products are highly differentiated an R&D-cartel under second-stage Cournot competition always enhances net total surplus. For values of β below $1/12$ the effect of an R&D-cartel on net total surplus depends on specific parameter values.

If competition increases because the number of firms in the industry expands, the competition-diminishing effect of allowing firms to form an industry wide R&D-cartel under second-stage Bertrand

competition features more prominently. That is, B_2 rises if n increases.[16] To put it differently, if the industry is large and pre-cooperative technological spillovers are small,[17] allowing for R&D-cartels under second-stage Bertrand competition will reduce net total surplus.

An industry wide RJV is likely to be socially desirable. Indeed, as observed by Arrow [1962], full sharing of information is socially beneficial given that the cost of transmitting knowledge is often close to zero. However, the desirability of an industry wide RJV depends on both the change in consumers' and producers' surplus a full dissemination of innovative knowledge induces. From the discussion in Section 1.5 it is clear that forcing technological spillovers to be maximal has a diluting effect on firms' incentives to invest in R&D, which possibly leads to a reduction in profits. Indeed, as the simulation results shown in Table 2D.1 of Appendix 2D.2 reveal, in some cases competitive profits fall if β increases. As a result, private and social incentives to form an industry-wide RJV could be in conflict.

Examining then the effects of full information sharing on net total surplus reveals that in most cases this leads to a net gain, as revealed by the simulation results presented in Table 2D.2 of Appendix 2D.2. Only if the industry is large and pre-cooperative spillovers are substantial would full information sharing be socially damaging. Indeed, under these circumstances any single firm's innovative knowledge would be disseminated among many competitors, inducing it to cut R&D expenses. And because pre-cooperative spillovers are already large, society's additional gain from full information sharing does not sufficiently compensate this drop in innovative activity. On the other hand, if products are independent, or if pre-cooperative technological spillovers are modest, an industry-wide RJV is always desired. In the former situation any firm is a local monopolist, making it immune to allegedly increased competition through full dissemination of knowledge, while in the latter case social gains associated with a substantially increased spillover rate outweigh the concomitant reduction in R&D activity.

Finally, if firms cooperate in their R&D, social and private incentives to share innovative information are in agreement, as shown in Lemma 2.6 (see Appendix 2B.6 for the proof).

Lemma 2.6

$\forall n, \beta, \theta$, the following holds

$$\frac{\partial \hat{W}_{IIC}(0)}{\partial \beta}, \frac{\partial \hat{W}_{IIB}(0)}{\partial \beta}, \frac{\partial W_{III}(0)}{\partial \beta} > 0.$$

---

[16] $\partial B_2/\partial n = \theta^3(1-\theta)/\{2+2\theta(n-2)-\theta^2(n-1)\}^2 > 0$.

[17] In the limit, 'small' means below $\lim_{n \to \infty} B_2 = \theta/(2-\theta)$.

Lemma 2.4 and Lemma 2.6 together imply that if firms are allowed to cooperate in their R&D (but remain competitors in the production stage), they always will engage in that type of R&D-cooperative which enhances net total surplus most. That is, if firms cooperate in R&D, they settle for an RJV-cartel, a type of R&D-cooperative governments also prefer to an R&D-cartel. Observe that this consistency between social and private incentives is independent of the type of product market competition.

Allowing firms to cooperate in their research thus seems an elegant way out of Arrow's [1962] trade off. Indeed, not only will firms fully exchange innovative information once they are allowed to jointly set their R&D investments, the concomitant level of R&D activity also exceeds that under a full competitive regime, as shown in Lemma 2.2. However, according to Lemma 2.3, an R&D-cooperative also wants to determine collectively prices prevailing in the subsequent product market. But this reduces net total surplus, since

$$\min_{n,\beta,\theta} \{ \hat{W}_{IIC}(0) - \hat{W}_{III}(0) \} > 0.78,$$

$$\min_{n,\beta,\theta} \{ \hat{W}_{IIB}(0) - \hat{W}_{III}(0) \} > 0.79.$$

Indeed, extending the R&D-cooperative agreement to the production stage always leads to a loss in net total surplus. Observe that this holds for any size of technological spillover. That is, extending an RJV-cartel to collusion in the production stage is also socially harmful.

Given then this contradiction between social and private incentives to collusion in production, the fundamental question arises as to the desirability of an R&D-cooperative. Comparing social welfare under the full competitive regime with that arising under an RJV-cartel with product market collusion reveals that in general the former exceeds the latter, as shown by the simulation results presented in Table 2D.3 of Appendix 2D.2. In particular, only if products are independent and the number of firms in the industry is small, or if pre-cooperative spillovers are absent and the industry is large, is the endogenous market outcome preferred to the full competitive regime. In all other cases it leads to a reduction in social welfare. Moreover, if the number of firms in the industry is large enough (for the simulations presented in the appendix this means $n \geq 129$), any type of R&D-cooperative with collusion in the product market is always inferior compared to the full competitive scenario.

To conclude this section, there are three observations to be made. First, allowing firms to form R&D-cartels can lead to increased R&D efforts if the combined-profits externality outweighs the competitive-advantage externality. That is, if research is complementary rather than substitutable. Indeed, under both second-stage Cournot and Bertrand competition, total R&D activity of an RJV-cartel exceeds competitive R&D efforts. These results are also reported by Kamien et al. [1992]. In addition, it is shown here that an industry wide monopoly with full sharing of innovative information leads to the highest level of R&D investments. Second, under both second-stage Cournot and Bertrand competition, participants in an R&D-cartel want to exchange innovative information as much as possible, a result Katz [1986] also reports for second-stage Cournot competition. As shown here, this information sharing incentive also holds if firms collude in the product market. Third, the endogenous

R&D-cooperative, an RJV-cartel, is also that which maximizes net total surplus. That is, social and private incentives towards R&D-cooperatives are in agreement. However, firms are tempted to extend this agreement to the production stage, an extension which is most likely to lower net social welfare vis-à-vis that emerging under a full competitive market system.

2.5 A case for subsidizing R&D?

Having outlined the effects of allowing for different types of R&D-cooperatives, the provision of optimal R&D-subsidies is considered. The first step is to investigate whether or not R&D-subsidies are positive. The following lemma, proved in Appendix 2B.7, shows that this is the case, that is, are actually subsidies rather than taxes.

Lemma 2.7

Part (i)

$\forall n, \beta, \theta$, the following holds

$$\dot{s}_{IB}, \dot{s}_{IIC}, \dot{s}_{IIB}, \dot{s}_{II} > 0.$$

Part (ii)

If $n=2$, or $n \geq 3 \wedge \beta \geq 1/12$, $\forall \theta$ the following holds

$$\dot{s}_{IC} > 0.$$

The first result stated in Lemma 2.7 is that, contrary to the findings of Spencer and Brander [1983] and Bagwell and Staiger [1992, 1994], in most cases optimal R&D-subsidies are positive, a result which is independent of the type of research considered (that is, complementary R&D versus substitutable R&D). This is in line with Spence [1984], who also reports consistently positive R&D-subsidies. Observe that both the present analysis and Spence [1984] do not consider international trade issues. It thus appears that optimal R&D-subsidies are especially effective in national markets.[18]

---

[18] It should be noted however that Spencer and Brander [1983] and Bagwell and Staiger [1992, 1994] do not consider consumers' surplus when evaluating the R&D-stimulating policies. Given that the provision of R&D subsidies (with a concomitant tax on sales) is especially beneficial for consumers, as will become clear below, the results reported in the international trade literature could be biased.

There are situations however in which the optimal R&D-subsidy could be negative, as stated in the second part of Lemma 2.7. Note that in the framework employed here, optimal R&D-subsidies induce private R&D levels to be that labelled second-best by Suzumura [1992], that is, the social planner's R&D investment *given* the second stage of the production process (this issue is explored in more detail in Chapter 3). For second-stage Cournot competition Suzumura [1992] shows that this second-best level of R&D investment is exceeded by competitive R&D efforts if technological spillovers are small.[19] In the model employed here a *sufficient* condition for this not to happen is that $\beta > 1/12$ (indeed, for smaller technological spillovers the optimal R&D-subsidy under second-stage Cournot competition in a fully competitive market could still be positive). Observe that for small spillovers and a sufficiently large industry ($n \geq 4$), Spence [1984] also reports negative R&D-subsidies in case of competitive R&D and second-stage Cournot competition.

The explanation for a possible overinvestment in R&D lies in the interaction of two opposite factors. On the one hand, competitive forces induce firms to invest in R&D in order secure market share in the product market. On the other hand, technological spillovers dilute R&D investments for the reasons discussed in Sections 1.5 and 2.4. The latter effect is weak if technological spillovers are small. Hence, firms tend to invest too much in R&D and the optimal R&D-subsidy is negative. But if the competitive-advantage externality becomes more important, the competitive forces are counterbalanced, with the result that for large enough spillovers competitive R&D efforts fall below those considered second-best.

On the other hand, in case of second-stage Bertrand competition firms never invest too much in R&D. That is, for this type of product market competition optimal R&D-subsidies are always positive. Indeed, under second-stage Bertrand competition more of the surplus generated by any firms' R&D is directed to consumers vis-à-vis product market competition over quantities. Hence, even if spillovers are small, competitive forces are not strong enough to induce firms to overinvest in R&D.

Suzumura [1992] also shows that an R&D-cartel (with competition in the product market) always invests less than the second-best level. Accordingly, the optimal R&D-subsidy is always positive. Lemma 2.7 states that for the framework employed here, this results also holds under second-stage Bertrand competition. Moreover, it is not affected by the fact that firms could cooperate in the product market, since also in that case the optimal R&D-subsidy is positive.[20]

Another salient feature of optimal R&D-subsidies, proved in Appendix 2B.8, is stated in the following lemma.

---

[19] Suzumura [1992] considers a more general framework than the one employed here to investigate R&D cooperatives. His conclusions only apply to some extent however, since his analysis is restricted to second stage Cournot competition, and collusion in the product market is not considered.

[20] Observe that social welfare under the third regime differs from that under the first two scenarios. Hence, the second-best levels of R&D are different. This difference is discussed in extensive detail in Section 2.7.

Lemma 2.8

$\forall n, \beta, \theta$, the following holds

$$\frac{\partial s_{IC}^*}{\partial \beta}, \frac{\partial s_{IB}^*}{\partial \beta}, \frac{\partial s_{IIC}^*}{\partial \beta}, \frac{\partial s_{IIB}^*}{\partial \beta}, \frac{\partial s_{III}^*}{\partial \beta} > 0.$$

That is, all optimal R&D-subsidies are increasing in the spillover rate.[21] This is due to two forces set in motion by increasing technological spillovers. First, as deliberated upon extensively in the previous section, the incentives to invest in R&D diminish if more information flows freely between firms. Second, as will become clear below, net total surplus is increasing in the spillover rate. That is, the higher are technological spillovers, the more a social planner would invest in R&D (d'Aspremont and Jacquemin's [1988] first-best level of R&D investment is also increasing in β). Indeed, the more freely information flows among firms, the more efficient are R&D investments, the more likely any additional unit of R&D, net of costs, will contribute to net total surplus. These two forces together then imply that the difference between the optimal (second best) level of R&D investments and those provided by any market configuration is increasing in the spillover rate. Hence, the optimal R&D-subsidies are increasing functions of β.

A more heuristic explanation would be that governments should subsidize more heavily those activities which use scarce factors most efficiently. And the efficiency of any firm's R&D efforts is increasing in the technological spillover rate, since more firms benefit from the innovator's activities. Hence, R&D-subsidies should be higher, the higher the technological spillover rate. From a practical point of view this means that generic research should be subsidized more heavily than applied research, since the former type of R&D is associated with much higher technological spillovers than the latter class of R&D (see e.g. Katz [1986]).

Investigating the effects of providing optimal R&D-subsidies and imposing a production tax accordingly leads to the following proposition, the proof of which is in Appendix 2C.1.

---

[21] Under second stage Cournot competition Spence [1984] reports similar findings. Lemma 2.8 shows that this feature carries over to a framework with differentiated products such that product market competition can also be over prices, and that it is not affected if firms tacitly collude in the production stage.

Proposition 2.1

Part (i)

$\forall n, \beta, \theta$, the following holds

$$x_{IB}^{\bullet} > \hat{x}_{IB}(0), \quad Q_{IB}^{\bullet} > \hat{Q}_{IB}(0), \quad W_{IB}^{\bullet} > \hat{W}_{IB}(0),$$

$$x_{IIC}^{\bullet} > \hat{x}_{IIC}(0), \quad Q_{IIC}^{\bullet} > \hat{Q}_{IIC}(0), \quad W_{IIC}^{\bullet} > \hat{W}_{IIC}(0),$$

$$x_{IIB}^{\bullet} > \hat{x}_{IIB}(0), \quad Q_{IIB}^{\bullet} > \hat{Q}_{IIB}(0), \quad W_{IIB}^{\bullet} > \hat{W}_{IIB}(0),$$

$$x_{III}^{\bullet} > \hat{x}_{III}(0), \quad Q_{III}^{\bullet} > \hat{Q}_{III}(0), \quad W_{III}^{\bullet} > \hat{W}_{III}(0).$$

If $n = 2$, or $n \geq 3 \wedge \beta \geq 1/12$, $\forall \theta$ the following holds

$$x_{IC}^{\bullet} > \hat{x}_{IC}(0), \quad Q_{IC}^{\bullet} > \hat{Q}_{IC}(0), \quad W_{IC}^{\bullet} > \hat{W}_{IC}(0).$$

Part (ii)

$\forall n, \beta, \theta$, the following holds

$$p_{IB}^{\bullet} < \hat{p}_{IB}(0), \quad \pi_{IB}^{\bullet} < \hat{\pi}_{IB}(0),$$

$$p_{IIC}^{\bullet} < \hat{p}_{IIC}(0), \quad \pi_{IIC}^{\bullet} < \hat{\pi}_{IIC}(0),$$

$$p_{IIB}^{\bullet} < \hat{p}_{IIB}(0), \quad \pi_{IIB}^{\bullet} < \hat{\pi}_{IIB}(0),$$

$$p_{III}^{\bullet} < \hat{p}_{III}(0), \quad \pi_{III}^{\bullet} < \hat{\pi}_{III}(0).$$

If $n = 2$, or $n \geq 3 \wedge \beta \geq 1/12$, $\forall \theta$ the following holds

$$p_{IC}^{\bullet} < \hat{p}_{IC}(0), \quad \pi_{IC}^{\bullet} < \hat{\pi}_{IC}(0).$$

According to Proposition 2.1, under second-stage Bertrand competition or in case of an industry-wide monopoly, irrespective of the size of technological spillovers and the extent to which products are differentiated, subsidizing R&D optimally, for which firms are taxed in the product market, increases the level of R&D, output and social welfare, but lowers prices and net profits. The same is true for the partially cooperative game with second-stage Cournot competition. The effects for the fully competitive regime when firms compete over quantities are only the same if the conditions under which the optimal R&D-subsidy is positive are met. If not, the effects of the R&D-subsidy on market performance depend on specific parameter configurations.

Considering the change in net total surplus it is thus almost always beneficial for authorities to shift resources from the product market to the innovation stage. Proposition 2.1 states that this effectively results in a shift from producers' surplus to consumers' surplus. Moreover, the increase in consumers' surplus in most cases outweighs the drop in producers' surplus.

Although private incentives towards R&D-subsidies (with a concomitant production tax) are the opposite of social incentives (as shown in Proposition 2.1), authorities can simply impose the production tax and subsidize R&D accordingly. Under those circumstances, firms' best responses are those given by the subgame perfect Nash-Cournot equilibria as computed in Section 2.3. Hence, authorities achieve their objective.

To summarize this section, some observations considering optimal R&D-subsidies can be stated. First, optimal R&D-subsidies are almost always positive. Hence, even if changes in market configurations, such as cooperative behaviour in the pre-competitive stage, can enhance R&D investments, the market still fails to provide some second-best level of R&D investments. Optimal R&D-subsidies on the other hand can effectively induce firms to invest that amount in R&D a social planner would desire, given the second stage of the production process. Second, the larger the gap between social and private incentives to invest in R&D, the larger the optimal R&D-subsidy will be. As a result, industries where technological spillovers are high (that is, industries where firms conduct generic research) should be subsidized more heavily than those industries in which the free flow of information is modest (that is, industries where mainly applied research is carried out). Third, providing optimal R&D-subsides and taxing sales accordingly results in a shift from producers' surplus to consumers' surplus, a shift which almost always results in an increase in net total surplus.

2.6 R&D-cooperatives versus optimal R&D-subsidies

Recall that one motivation to consider R&D-subsidies and R&D-cooperatives within the same analytical framework is that it allows for a comparison of the two policies. Indeed, if either policy is to be implemented, some insight as to the relative effectiveness of each instrument is desired. The following proposition provides this comparison. The proof can be found in Appendix 2C.2.

Proposition 2.2

$\forall n, \beta, \theta$, the following holds

Part (i)

$$x_{IC}^* > \hat{x}_{IC}(0)|_{\beta=1}, \quad x_{IB}^* > \hat{x}_{IB}(0)|_{\beta=1}.$$

Part (ii)

$$x_{IC}^{\bullet} > \hat{x}_{IIC}(0), \quad x_{IB}^{\bullet} > \hat{x}_{IIB}(0).$$

Part (iii)

$$x_{IC}^{\bullet} > \hat{x}_{IIC}(0)|_{\beta=1} \quad \Leftrightarrow \quad b\gamma[2+\theta(n-1)]^2 > \frac{2n(n-1)(1-\beta)[3+\theta(n-1)][1+\beta(n-1)]}{[3+\theta(n-1)][1+\beta(n-1)]-2n},$$

$$x_{IB}^{\bullet} > \hat{x}_{IIB}(0)|_{\beta=1} \quad \Leftrightarrow \quad b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 >$$

$$\frac{2n(n-1)(1-\theta)(1-\beta)[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]}{[3+\theta(n-4)][1+\beta(n-1)]-2n(1-\theta)}.$$

The first part of proposition 2.2 states that providing firms with an optimal R&D-subsidy is more effective in promoting private R&D investments than allowing for an industry wide RJV. This observation is readily explained. As discussed in Sections 1.5 and 2.4, setting the competitive-advantage externality at its maximum results in the lowest level of private R&D spending under both second-stage Cournot and Bertrand competition. On the other hand, any firms' optimal R&D investment is an increasing function of the optimal R&D-subsidy (see Section 2.3). That is, the provision of (positive) R&D-subsidies always promotes private R&D spending. Hence, subsidized competitive R&D efforts exceed those of a non-subsidized RJV. Notice that this result holds for all values any parameter of the model is allowed to take. That is, even if the optimal R&D-subsidy is negative (see Lemma 2.7) a non-subsidized RJV will conduct less R&D compared to the subsidized competitive alternative.

More important however is that providing optimal R&D-subsidies is more effective in raising private R&D spending than allowing firms to form an industry-wide R&D-cartel. That is, changing the market structure by imposing the tax/subsidy regime on the industry results in a greater effect than allowing firms to cooperatively determine their R&D efforts. Recall that the obvious advantages of subsidizing R&D over any R&D-cooperative are that negative pecuniary externalities will not be internalized, and that the threat of tacit collusion in the product market is not enhanced. In addition, according to Proposition 2.2, providing optimal R&D-subsidies result a larger increase in private R&D spending than allowing for an R&D-cartel.

However, as derived in Section 2.4, the endogenous market outcome without optimal R&D-subsidies will be an industry wide RJV-cartel. And this type of R&D-cooperative is also most desired from a social welfare point of view. As shown in Proposition 2.2, $b\gamma$ has to be 'large enough' for the provision of optimal R&D-subsidies also to beat the formation of an industry wide RJV-cartel in terms of R&D investment promotion. It can be shown though that no second-order condition, stability

condition, or the conditions for the tax rate to be below 1, are sufficient for those stated in Proposition 2.2 to hold. On the other hand, given that the conditions in the proposition are of the form $b\gamma > \alpha(n,\beta,\theta)$, no condition coming from the model impedes these to be met. On the contrary, examples are readily derived in which optimally subsidized competitive R&D exceeds that of an RJV-cartel.

In short, optimally subsidizing noncooperative R&D is more effective in promoting private R&D investment than allowing for an RJV or R&D-cartel, and, in most cases, is also more effective than allowing for an RJV-cartel. If either of the policies is to be implemented, allowing firms to cooperate in their R&D is not the most effective direction to follow. In practise however, both instruments are used simultaneously, a policy portfolio that is to be analyzed next.

2.7 A case for subsidized cooperation?

A second motivation to consider R&D-cooperatives and R&D-subsidies within the same framework is the analysis of subsidized R&D-cooperatives. As mentioned in Section 1.7, there is very little literature on this issue. Indeed, only Spence [1984] briefly considers this policy portfolio.

First, the optimal R&D-subsidies are to be compared. That is, should an R&D-cooperative be subsidized more or less heavily than a competitive R&D market? The following lemma, the proof of which is given in Appendix 2B.9, clarifies this.

Lemma 2.9

$\forall n, \beta, \theta$, the following holds

$$\dot{s}_{IC} < \dot{s}_{IIC} \quad \Leftrightarrow \quad \beta \in [0, C_2),$$

$$\dot{s}_{IC} > \dot{s}_{IIC} \quad \Leftrightarrow \quad \beta \in (C_2, 1],$$

and

$$\dot{s}_{IB} < \dot{s}_{IIB} \quad \Leftrightarrow \quad \beta \in [0, B_2),$$

$$\dot{s}_{IB} > \dot{s}_{IIB} \quad \Leftrightarrow \quad \beta \in (B_2, 1],$$

where C_2 and B_2 are defined in Lemma 2.1.

According to Lemma 2.9, whenever competitive R&D efforts exceed that of an R&D-cartel, the reverse is true for the optimal R&D-subsidy, and vice versa. This observations has two consequences. First, recall from the discussion in Section 2.4 that if research is complementary, an R&D-cartel

devotes more resources to R&D vis-à-vis competitive R&D. Lemma 2.9 then implies that competitive R&D should be subsidized more heavily than cooperative R&D when research is complementary. According to the same logic, an R&D-cartel should be subsidized more heavily than competitive R&D when research is substitutable.

Second, providing the optimal R&D-subsidy leads to a convergence in resources devoted to R&D. Comparing then the first two regimes under the provision of optimal R&D-subsidies leads to the following remarkable observation (see Tables 2A.1 through 2A.4 of Appendix 2A).

Proposition 2.3

$\forall n, \beta, \theta$, the following holds

$$x_{IJ}^{\bullet} = x_{IIJ}^{\bullet}, \quad Q_{IJ}^{\bullet} = Q_{IIJ}^{\bullet}, \quad p_{IJ}^{\bullet} = p_{IIJ}^{\bullet}, \quad \pi_{IJ}^{\bullet} = \pi_{IIJ}^{\bullet}, \quad W_{IJ}^{\bullet} = W_{IIJ}^{\bullet},$$

for $J = C, B$.

Proposition 2.3 states that encouraging private R&D investments by allowing firms to participate in R&D-cartels *and in addition* optimally subsidizing this agreement has the same effect on market performance and social welfare as optimally subsidizing noncooperative R&D. Since Proposition 2.3 holds for all levels of technological spillover, it is in particular valid for $\beta = 1$. That is, subsidizing an RJV optimally leads to the same market outcome and social welfare as optimally subsidizing an RJV-cartel.

According to Proposition 2.1 it is always beneficial for net total surplus to implement the subsidy/tax policy. In addition, Proposition 2.2 states that providing optimal R&D-subsidies is in most cases more effective in promoting private R&D efforts than allowing for any of the R&D-cooperatives considered. And Proposition 2.3 reveals that optimally subsidized competitive R&D yields exactly the same market outcome as an optimally subsidized R&D-cartel. Indeed, these three observations together imply that if the provision of optimal R&D-subsidies with a concomitant tax on sales is a feasible instrument to stimulate private R&D spending, allowing firms to cooperatively determine their R&D investment is a redundant policy tool.

The only two things left to consider then, if optimal R&D-subsidies are provided, is the effects on R&D investment, profits and net total surplus of allowing firms to cooperate in both stages of the production process, and the consequences of full information sharing on competitive markets and within a monopoly.

2.7.1 Subsidized research and development

The following lemma, which is proved in Appendix 2B.10, summarizes the comparison of optimally subsidized R&D levels.

Lemma 2.10

Part (i)

$\forall n, \beta, \theta$, the following holds

$$x_{IC}^{\bullet} > x_{III}^{\bullet},$$

$$x_{IB}^{\bullet} > x_{III}^{\bullet}.$$

Part (ii)

$\forall n, \beta, \theta$, the following holds

$$x_{IC}^{\bullet}|_{\beta=1} > x_{IC}^{\bullet}, x_{III}^{\bullet},$$

$$x_{IB}^{\bullet}|_{\beta=1} > x_{IB}^{\bullet}, x_{III}^{\bullet}.$$

The first part of Lemma 2.10 reveals that tacit collusion in the product market always leads to a reduction in private R&D efforts if research is optimally subsidized. As shown by the simulation results presented in Table 2D.4 of Appendix 2D.2, competing firms generally receive a higher optimal R&D-subsidy than a monopoly.[22] According to Lemma 2.10, this higher subsidy outweighs the rent capturing effect of collusion in the product market.

The second part of Lemma 2.10 shows that subsidized competitive R&D efforts are increasing in the spillover rate, a result quite opposite to that implied by Lemma 2.2. The reason for this is that on the one hand incentives to invest in R&D decline whenever β increases. But on the other hand, as shown in Lemma 2.8, the optimal R&D-subsidies are higher, the larger are technological spillovers. According to Lemma 2.10 then, the higher R&D-subsidies more than offset the increased competitive-advantage externality.

The highest level of private R&D spending is thus achieved by subsidizing competing firms optimally, while encouraging them to fully share the results of their R&D. That is, an optimally subsidized RJV (or RJV-cartel) yields the most R&D intensive industry. Whether or not this market configuration can be sustained by private incentives is to be considered next.

---

[22] Both under second stage Cournot and Bertrand competition, only if β is close to zero it could be that $s_I^{\bullet} < s_{III}^{\bullet}$. In that case, the difference between the second-best level of R&D investment and that provided my the market is smaller under a fully competitive regime than in case of a monopoly. It can be shown that the optimal R&D subsidy given to an R&D cooperative always exceed that provided to a monopoly.

2.7.2 Private incentives towards subsidized R&D

If firms receive the optimal R&D-subsidy and are confronted with a concomitant tax on their sales, they are still inclined to cooperate in as many stages as possible, as shown in the following lemma (the proof can be found in Appendix 2B.11).

Lemma 2.11

Part (i)

$\forall n, \beta, \theta$, the following holds

$$\pi_{IC}^* < \pi_{III}^*,$$

$$\pi_{IB}^* < \pi_{III}^*.$$

Part (ii)

$\forall n, \beta, \theta$, the following holds

$$\frac{\partial \pi_{IC}^*}{\partial \beta}, \frac{\partial \pi_{IB}^*}{\partial \beta} < 0,$$

$$\frac{\partial \pi_{III}^*}{\partial \beta} > 0.$$

According to Lemma 2.11, the endogenous market outcome when optimal R&D-subsidies are provided is an industry-wide RJV-cartel with collusion in the product market. Indeed, this is the same market structure as that emerging when no R&D-subsidies are provided. However, optimal R&D-subsidies have an effect on private incentives to share information if collusion in the product market is effectively prohibited. That is, whether firms cooperate in R&D or not, they do not want to share innovative information if competition in the second stage of the production process is enforced. Hence, an optimally subsidized RJV will not be the endogenous market outcome if firms are prohibited to jointly set production or prices. In that case rather, each firm tries to dampen technological leakage as much as possible, independent of whether it determines its R&D investments jointly are unilaterally.

2.7.3 Social incentives towards subsidized R&D

As proved in Appendix 2B.12, net total surplus when optimal R&D-subsidies are provided is ranked as follows (with concomitant social incentives towards information sharing)

Lemma 2.12

Part (i)

$\forall n, \beta, \theta$, the following holds

$$W_{IC}^* > W_{III}^*,$$

$$W_{IB}^* > W_{III}^*.$$

Part (ii)

$\forall n, \beta, \theta$, the following holds

$$\frac{\partial W_{IC}^*}{\partial \beta}, \frac{\partial W_{IB}^*}{\partial \beta}, \frac{\partial W_{III}^*}{\partial \beta} > 0.$$

Indeed, an (optimally subsidized) industry-wide monopoly always yields a lower net total surplus compared to the (optimally subsidized) competitive alternative. Again the redundancy of allowing for the formation of an R&D-cartel is revealed. On the other hand, in case optimal R&D-subsidies are provided, under all regimes considered it is socially beneficial if firms exchange fully innovative information, as shown by the second part of Lemma 2.12.

2.8 An optimal R&D stimulating policy

The policy implications of Lemma 2.12 are clear: firms should be encouraged to form an industry wide RJV and this agreement should be optimally subsidized. Considering all alternatives, this policy has a number of advantages. First, it does not facilitate collusive behaviour in the product market. Second, the possibility that firms collectively decide to cut R&D investment is absent. Third, it will yield the most R&D-intensive industry. Moreover, given that firms are taxed for the R&D-subsidy in the product market, there is a shift from producers' surplus to consumers' surplus that leads to a gain in net total surplus.

The proposal to optimally subsidize an RJV will meet with opposition from the industry. First, as shown in Proposition 2.1, firms are never in favour of the R&D-subsidy/production-tax policy since it always leads to a reduction in profits. However, as discussed in Section 2.5, if firms are confronted with the production tax, their optimal behaviour is that described in Section 2.3 and authorities achieved their objective. Second, and more importantly, if firms receive the optimal R&D-subsidy while not being allowed to cooperate in any stage of the production process, they do not have any incentive to share the fruits of their research. Indeed, the trade-off identified by Arrow [1962] is still features prominently. A solution to this second problem could be that firms are always taxed in the product market, but receive the R&D-subsidy *only if* they fully share the results of their R&D.

Finally observe that optimally subsidizing an RJV also yields higher net total surplus than allowing for a non-subsidized RJV-cartel (the optimal R&D stimulating policy when R&D-subsidies are not provided). Indeed, subsidizing an RJV-cartel optimally leads to an increase in net total surplus (Proposition 2.1). But this yields exactly the same market outcome as subsidizing an RJV (Proposition 2.3). The price to be paid, of course, is that private incentives to form an RJV-cartel coincide with social incentives, whereas firms will object to the subsidization of competitive R&D (with a concomitant sales tax) while being forced to exchange fully the fruits of their research.

2.9 Conclusions

In this chapter one of the first analyses is presented considering two widely used R&D stimulating policies within the same analytical framework: allowing for R&D-cooperatives and providing R&D-subsidies. From the discussion several conclusions can be drawn.

As shown here (and by several studies published over the last decade; see Chapter 1), cooperative R&D can lead to increased R&D efforts. In the model employed here this happens when research is complementary. In case of substitutable research on the other hand, an R&D-cooperative will decide to devote less resources to R&D vis-à-vis a competitive market. Also, it is shown that cooperative R&D is more likely to exceed noncooperative R&D the more distinct products are. In that case the competitive-advantage externality features less prominently while the combined-profits externality is likely to be positive.

In absence of optimal R&D-subsidies, under both second-stage Cournot and Bertrand competition, the endogenous market outcome is an RJV-cartel, if collusion in the product market is effectively prohibited. This market outcome also yields the highest level of net total surplus. However, the threat of firms extending the R&D-cooperative agreement to the production stage is obvious, since this always yields higher profits. In most cases, net total surplus will then be below that of a fully competitive market.

In most cases, optimal R&D-subsidies are positive. This result differs from those reported in the international trade literature (see Spencer and Brander [1983] and Bagwell and Staiger [1992, 1994]) but is also reported by Spence [1984], who also only considers a domestic market. As to the size of optimal R&D-subsidies it is shown that generic research should be subsidized more heavily

than applied research, since the former type of R&D is associated with higher spillover rates. Also, if research is complementary, competitive firms should receive a higher R&D-subsidy than cooperating firms, while the reverse holds for substitutable research.

The general effect of subsidizing R&D optimally, while taxing firms for it in the product market, is a shift from producers' surplus to consumers' surplus. Under most circumstances this shift leads to a gain in net total surplus.

Comparing the provision of optimal R&D-subsidies with allowing for an R&D-cooperative, reveals that the former policy in most cases is more effective than the latter to promote private R&D investments. In particular, providing a competitive market with an optimal R&D-subsidy yields higher R&D investments than allowing for an RJV or R&D-cartel, while in most cases it also produces greater effect than sustaining an RJV-cartel.

A remarkable observation is that encouraging private R&D investments by allowing firms to participate in R&D-cartels *and in addition* optimally subsidizing this agreement has the same effect on market performance and social welfare as subsidizing optimally noncooperative R&D. Together with the observations mentioned before this implies that allowing firms to cooperate in their R&D is a redundant policy. Rather, the optimal R&D-stimulating policy is to optimally subsidize competitive R&D and to encourage firms to fully exchange innovative information, that is, to optimally subsidize an RJV.

Although the analysis presented here is one of the first in the area, it should be borne in mind that only a restricted framework is considered. Hence, all conclusions drawn should be treated with caution. However, in the next chapter it will be shown that some key results presented thus far carry over to a more general framework.

Appendix 2A Market equilibria

Table 2A.1 Equilibrium without and with an R&D-subsidy; Cournot competition; No cooperation in R&D, no cooperation in output

$\hat{x}_{IC}(0)$

$$\frac{2(a-A)[2+\theta\{(n-2)-\beta(n-1)\}]}{b\gamma(2-\theta)[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]}$$

$\hat{p}_{IC}(0)$

$$a - \frac{b\gamma(a-A)(2-\theta)[2+\theta(n-1)][1+\theta(n-1)]}{b\gamma(2-\theta)[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]}$$

$\hat{Q}_{IC}(0)$

$$\frac{\gamma n(a-A)(2-\theta)[2+\theta(n-1)]}{b\gamma(2-\theta)[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]}$$

$\hat{\pi}_{IC}(0)$

$$\frac{\gamma(a-A)^2\{b\gamma(2-\theta)^2[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}]^2\}}{\{b\gamma(2-\theta)[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]\}^2}$$

$\hat{W}_{IC}(0)$

$$\frac{n\gamma(a-A)^2\{b\gamma(2-\theta)^2[3+\theta(n-1)][2+\theta(n-1)]^2-4[2+\theta\{(n-2)-\beta(n-1)\}]^2\}}{2\{b\gamma(2-\theta)[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]\}^2}$$

s_{IC}

$$\frac{\gamma(a-A)\{(2-\theta)[3+\theta(n-1)][1+\beta(n-1)]-2[2+\theta\{(n-2)-\beta(n-1)\}]\}}{(2-\theta)\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}}$$

t_{IC} [a]

$$2[1+\beta(n-1)][3+\theta(n-1)]$$
$$\times \frac{(2-\theta)[3+\theta(n-1)][1+\beta(n-1)]-2[2+\theta\{(n-2)-\beta(n-1)\}]}{2b\gamma(2-\theta)[2+\theta(n-1)]^2-[1+\beta(n-1)][3+\theta(n-1)]\alpha_6^C}$$

x_{IC}^{*}

$$\frac{(a-A)[3+\theta(n-1)][1+\beta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}$$

p_{IC}^{*}

$$a - \frac{b\gamma(a-A)(2-\theta)[2+\theta(n-1)][1+\theta(n-1)]}{b\gamma(2-\theta)[2+\theta(n-1)^2]-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]}$$

Q_{IC}^{*}

$$\frac{\gamma n(a-A)[2+\theta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}$$

π_{IC}^{*}

$$\frac{\gamma(a-A)^2\{2b\gamma[2+\theta(n-1)]^2-[1+\beta(n-1)]^2[3+\theta(n-1)]^2\}}{2\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2}$$

W_{IC}^{*}

$$\frac{\gamma n(a-A)^2[3+\theta(n-1)]}{2\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}}$$

[a] $\alpha_6^C = 2+\theta(2n-3)+\theta^2(n-1)-\beta(n-1)[6+\theta(2n-1)-\theta^2(n-1)]$.

Table 2A.2   **Equilibrium with and without an R&D-subsidy; Cournot competition; Cooperation in R&D, no cooperation in output**

$\hat{x}_{IIC}(0)$

$$\frac{2(a-A)[1+\beta(n-1)]}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2}$$

$\hat{p}_{IIC}(0)$

$$a-\frac{b\gamma(a-A)[2+\theta(n-1)][1+\theta(n-1)]}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2}$$

$\hat{Q}_{IIC}(0)$

$$\frac{\gamma n(a-A)[2+\theta(n-1)]}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2}$$

$\hat{\pi}_{IIC}(0)$

$$\frac{\gamma(a-A)^2}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2}$$

$\hat{W}_{IIC}(0)$

$$\frac{n\gamma(a-A)^2\{b\gamma[3+\theta(n-1)][2+\theta(n-1)]^2-4\{1+\beta(n-1)]^2\}}{2\{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2\}^2}$$

$s_{IIC}^{\bullet}$

$$\frac{\gamma(a-A)[1+\beta(n-1)][1+\theta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}$$

$t_{IIC}^{\bullet}$

$$\frac{2[1+\theta(n-1)][3+\theta(n-1)][1+\beta(n-1)]^2}{2b\gamma[2+\theta(n-1)]^2-[1-\theta(n-1)][3+\theta(n-1)][1+\beta(n-1)]^2}$$

$x_{IIC}^{\bullet}$

$$\frac{(a-A)[3+\theta(n-1)][1+\beta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+(n-1)\beta]^2}$$

$p_{IIC}^{\bullet}$

$$a-\frac{b\gamma(a-A)(2-\theta)[2+\theta(n-1)][1+\theta(n-1)]}{b\gamma(2-\theta)[2+\theta(n-1)^2]-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]}$$

$Q_{IIC}^{\bullet}$

$$\frac{\gamma n(a-A)[2+\theta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}$$

$\pi_{IIC}^{\bullet}$

$$\frac{\gamma(a-A)^2\{2b\gamma[2+\theta(n-1)]^2-[1+\beta(n-1)]^2[3+\theta(n-1)]^2\}}{2\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2}$$

$W_{IIC}^{\bullet}$

$$\frac{\gamma n(a-A)^2[3+\theta(n-1)]}{2\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}}$$

Table 2A.3 Equilibrium without and with an R&D-subsidy; Bertrand competition; No cooperation in R&D, no cooperation in output

$$\hat{x}_{IB}(0) \qquad \frac{2\Delta(a-A)[1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]}$$

$$\hat{p}_{IB}(0) \qquad a - \frac{b\gamma(a-A)[1+\theta(n-1)][1+\theta(n-2)][2+\theta(2n-3)][2+\theta(n-3)]}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]}$$

$$\hat{Q}_{IB}(0) \qquad \frac{\gamma n(a-A)[1+\theta(n-2)][2+\theta(2n-3)][2+\theta(n-3)]}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]}$$

$$\hat{\pi}_{IB}(0) \qquad \frac{\gamma(a-A)^2[1+\theta(n-2)]\{b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(2n-3)]^2[2+\theta(n-3)]^2-2\Delta^2[1+\theta(n-2)]\}}{\{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]\}^2}$$

$$\hat{W}_{IB}(0) \qquad \gamma n(a-A)^2[1+\theta(n-2)]$$

$$\times \frac{b\gamma[3+\theta(n-4)][1+\theta(n-1)][2+\theta(2n-3)]^2[2+\theta(n-3)]^2-4\Delta^2[1+\theta(n-2)]}{2\{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]\}^2}$$

$$s_{IB} \qquad \frac{\gamma(a-A)[1+\theta(n-2)]\{[2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-2\Delta\}}{[2+\theta(2n-3)]\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}}$$

$$t_{IB}^{\;b} \qquad \frac{2[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]\{[2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-2\Delta\}}{2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2+[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]\alpha_6^B}$$

$$x_{IB}^{\bullet} \qquad \frac{(a-A)[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$p_{IB}^{\bullet} \qquad a - \frac{b\gamma(a-A)[1+\theta(n-1)][2+\theta(n-3)][1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$Q_{IB}^{\bullet} \qquad \frac{\gamma n(a-A)[2+\theta(n-3)][1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$\pi_{IB}^{\bullet} \qquad \gamma(a-A)^2[1+\theta(n-2)]$$

$$\times \frac{2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)]^2[1+\beta(n-1)]^2}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2}$$

$$W_{IB}^{\bullet} \qquad \frac{\gamma n(a-A)^2[1+\theta(n-2)][3+\theta(n-4)]}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}}$$

---

a $\Delta = 2+3\theta(n-2)+\theta^2[(n-1)(n-2)-(2n-3)]-\theta\beta(n-1)[1+\theta(n-2)]$.

b $\alpha_6^B = [2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-4\Delta$.

Table 2A.4   **Equilibrium with and without an R&D-subsidy; Bertrand competition; Cooperation in R&D, no cooperation in output**

$$\hat{x}_{IIB}(0) \qquad \frac{2(a-A)(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}$$

$$\hat{p}_{IIB}(0) \qquad a-\frac{b\gamma(a-A)[1+\theta(n-1)][1+\theta(n-2)][2+\theta(n-3)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}$$

$$\hat{Q}_{IIB}(0) \qquad \frac{n\gamma(a-A)[1+\theta(n-2)][2+\theta(n-3)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}$$

$$\hat{\pi}_{IIB}(0) \qquad \frac{\gamma(a-A)^2(1-\theta)[1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}$$

$$\hat{W}_{IIB}(0) \qquad n\gamma(a-A)^2[1+\theta(n-2)]$$
$$\times \frac{b\gamma[1+\theta(n-1)][3+\theta(n-4)][2+\theta(n-3)]^2-4(1-\theta)^2[1+\theta(n-2)][1+\beta(n-1)]^2}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2\}^2}$$

$$s^{*}_{IIB} \qquad \frac{\gamma(a-A)[1+\theta(n-2)]^2[1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$t^{*}_{IIB} \qquad \frac{2[3+\theta(n-4)][1+\theta(n-2)]^2[1+\beta(n-1)]^2}{2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2+(\theta n-1)[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$x^{*}_{IIB} \qquad \frac{(a-A)[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$p^{*}_{IIB} \qquad a-\frac{b\gamma(a-A)[1+\theta(n-1)][2+\theta(n-3)][1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$Q^{*}_{IIB} \qquad \frac{\gamma n(a-A)[2+\theta(n-3)][1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$\pi^{*}_{IIB} \qquad \gamma(a-A)^2[1+\theta(n-2)]$$
$$\times \frac{2b\gamma(1-\theta)[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)]^2[1+\beta(n-1)]^2}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2}$$

$$W^{*}_{IIB} \qquad \frac{\gamma n(a-A)^2[1+\theta(n-2)][3+\theta(n-4)]}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}}$$

Table 2A.5 Equilibrium with and without an R&D-subsidy;
Cooperation in R&D, cooperation in output

$\hat{x}_{III}(0)$

$$\frac{(a-A)[1+\beta(n-1)]}{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2}$$

$\hat{p}_{III}(0)$

$$a-\frac{b\gamma(a-A)[1+\theta(n-1)]}{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2}$$

$\hat{Q}_{III}(0)$

$$\frac{\gamma n(a-A)}{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2}$$

$\hat{\pi}_{III}(0)$

$$\frac{\gamma(a-A)^2}{2\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}}$$

$\hat{W}_{III}(0)$

$$\frac{n\gamma(a-A)^2\{3b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}}{2\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}^2}$$

$\overset{\bullet}{s}_{III}$

$$\frac{\gamma(a-A)[1+\beta(n-1)]}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}$$

$\overset{\bullet}{t}_{III}$

$$\frac{6[1+\beta(n-1)]^2}{8b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}$$

$\overset{\bullet}{x}_{III}$

$$\frac{3(a-A)[1+\beta(n-1)]}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}$$

$\overset{\bullet}{p}_{III}$

$$a-\frac{2b\gamma(a-A)[1+\theta(n-1)]}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}$$

$\overset{\bullet}{Q}_{III}$

$$\frac{2\gamma n(a-A)}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2}$$

$\overset{\bullet}{\pi}_{III}$

$$\frac{\gamma(a-A)^2\{8b\gamma[1+\theta(n-1)]-9[1+\beta(n-1)]^2\}}{2\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}^2}$$

$\overset{\bullet}{W}_{III}$

$$\frac{3\gamma n(a-A)^2}{2\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}}$$

Appendix 2B Proofs of Lemmata

2B.1 Proof of Lemma 2.1

Comparing the different levels of non-subsidized R&D leads to

$$\hat{x}_{IC}(0) - \hat{x}_{IIC}(0) = \frac{2b\gamma(a-A)(n-1)[2+\theta(n-1)]^2}{b\gamma(2-\theta)[2+\theta(n-1)]^2 - 2\{[2+\theta(n-2)]-\theta\beta(n-1)\}[1+\beta(n-1)]}$$

$$\times \frac{\theta-2\beta}{b\gamma[2+\theta(n-1)]^2 - 2[1+\beta(n-1)]^2},$$

$$\hat{x}_{IC}(0) - \hat{x}_{III}(0) = \frac{b\gamma(a-A)(n-1)}{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2}$$

$$\times \frac{\beta[8+8\theta(n-1)+2\theta^2(n-1)^2-\theta^3(n-1)^3]-\theta[4+2\theta(n-1)+\theta^2(n-1)]}{b\gamma(2-\theta)[2+\theta(n-1)]^2 - 2\{[2+\theta(n-2)]-\theta\beta(n-1)\}[1+\beta(n-1)]}.$$

$$\hat{x}_{IIC}(0) - \hat{x}_{III}(0) = \frac{b\gamma(a-A)(n-1)^2\theta^2[1+\beta(n-1)]}{\{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2\}\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}},$$

and

$$\hat{x}_{IB}(0) - \hat{x}_{IIB}(0) = \frac{2b\gamma(a-A)(n-1)[1+\theta(n-1)][1+\theta(n-2)][2+\theta(n-3)]^2}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 - 2\Delta[1+\theta(n-2)][1+\beta(n-1)]}$$

$$\times \frac{\theta[1+\theta(n-2)]-\beta[2+2\theta(n-2)-\theta^2(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - 2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2},$$

$$\hat{x}_{IB}(0) - \hat{x}_{III}(0) = \frac{b\gamma(a-A)(n-1)[1+\theta(n-1)]}{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2}$$

$$\times \frac{\beta[8+16\theta(n-2)+2\theta^2[5n^2-20n+19]+\theta^3(n-1)(2n^2-9n+11)]-\theta[4+2\theta(3n-7)+\theta^2(2n^2-11n+13)]}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 - 2\Delta[1+\theta(n-2)][1+\beta(n-1)]}.$$

$$\hat{x}_{III}(0) - \hat{x}_{IIB}(0) = \frac{b\gamma(a-A)\theta^2(n-1)^2}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - 2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}$$

$$\times \frac{[1+\theta(n-1)][1+\beta(n-1)]}{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2}.$$

Then observe that

$$x_{IC}(0) - x_{IIC}(0) > 0 \quad \Leftrightarrow \quad \beta < C_2,$$

$$x_{IC}(0) - x_{III}(0) > 0 \quad \Leftrightarrow \quad \beta < C_1,$$

$$x_{III}(0) - x_{IIC}(0) > 0 \quad \Leftrightarrow \quad \forall \beta \in [0,1],$$

and

$$x_{IB}(0) - x_{IIB}(0) > 0 \quad \Leftrightarrow \quad \beta < B_2,$$

$$x_{IB}(0) - x_{III}(0) > 0 \quad \Leftrightarrow \quad \beta < B_1,$$

$$x_{III}(0) - x_{IIB}(0) > 0 \quad \Leftrightarrow \quad \forall \beta \in [0,1].$$

Finally, it is to be examined whether or not the boundaries on β, for which the inequalities stated in Lemma 2.1 hold, do not cross. Observe that under second-stage Cournot competition

$$C_2 - C_1 = \frac{\theta^2(n-1)\{4 + 2\theta(n-2) - \theta^2(n-1)\}}{2\{8[1 + \theta(n-1)] + \theta^2(n-1)(2-\theta)\}},$$

the denominator of which is clearly positive. The numerator is positive *iff* $-2/(n-1) < \theta < 2$, and $\theta \neq 0$, conditions which are all satisfied under the parameter space considered in the analysis. Further, under second-stage Bertrand competition

$$B_2 - B_1 = \frac{\theta^2(n-1)\{4 + 2\theta(5n-8) + \theta^2[8n^2 - 27n + 21] + \theta^3[2n^3 - 11n^2 + 18n - 9]\}}{\{2(1-\theta) + \theta(n-1)(2-\theta)\}\{16\theta(n-2) + 2\{4 + \theta^2[5n^2 - 20n + 19]\} + \theta^3(n-1)[2n^2 - 9n + 11]\}},$$

the denominator of which also is obviously positive. The same holds for the numerator if it is observed that it is increasing in n and positive for $n = 2$. Hence, both under second-stage Cournot and Bertrand competition the regions in terms of β as stated in Lemma 2.1 are well defined.

Q.E.D.

2B.2 Proof of Lemma 2.2

Part (i)

First observe that $\forall n, \beta, \theta$,

$$\hat{x}_{IIC}(0)|_{\beta=1} - \hat{x}_{IC}(0) = \frac{2(a-A)(n-1)}{b\gamma[2 + \theta(n-1)]^2 - 2n^2}$$

$$\times \frac{b\gamma[2 + \theta(n-1)]^2[2 - 2\theta + \theta\beta] + 2n(1-\beta)[2 + \theta\{(n-2) - \beta(n-1)\}]}{b\gamma(2-\theta)[2 + \theta(n-1)]^2 - 2[2 + \theta\{(n-2) - \beta(n-1)\}][1 + \beta(n-1)]} > 0,$$

and

$$\hat{x}_{IIB}(0)\big|_{\beta=1}-\hat{x}_{IB}(0)=\frac{2(a-A)(n-1)[1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)n^2[1+\theta(n-2)]}$$

$$\times\frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2\{2+2\theta(n-3)-\theta^2(3n-5)+\theta\beta[1+\theta(n-2)]\}}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]}.$$

For the second of these to be positive, it suffice that

$$2+\theta\{2(n-3)-\theta(3n-5)\}+\theta\beta[1+\theta(n-2)]$$

or

$$2+\theta\{2(n-3)-\theta(3n-5)\}$$

is positive. The latter expression has two roots in θ, one positive and one negative. For θ below the positive root, that is, for

$$\theta<\frac{n-3+\sqrt{(n-1)(n+1)}}{3n-5},$$

it is true that $\hat{x}_{IIB}(0)\big|_{\beta=1}-\hat{x}_{IB}(0)>0$. Observe that the positive root is decreasing in n and that

$$\lim_{n\to\infty}\left\{\frac{n-3+\sqrt{(n-1)(n+1)}}{3n-5}\right\}=\frac{2}{3}.$$

The proof of the first part of the lemma is complete if it is realized that $\forall n,\beta,\theta$,

$$\frac{\partial\hat{x}_{IIC}(0)}{\partial\beta}=\frac{2(n-1)(a-A)\{b\gamma[2+\theta(n-1)]^2+2[1+\beta(n-1)]^2\}}{\{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2\}^2}>0,$$

$$\frac{\partial\hat{x}_{IIB}(0)}{\partial\beta}=2(n-1)(a-A)(1-\theta)[1+\theta(n-2)]$$

$$\times\frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2+2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2\}^2}>0.$$

Part (ii)

First observe that $\forall n,\beta,\theta$,

$$\hat{x}_{III}(0)\big|_{\beta=1}-\hat{x}_{IIC}(0)\big|_{\beta=1}=\frac{n(n-1)^2b\gamma\theta^2(a-A)}{\{2b\gamma[1+\theta(n-1)]-n^2\}\{b\gamma[2+\theta(n-1)]^2-2n^2\}}>0,$$

and

$$\hat{x}_{III}(0)\big|_{\beta=1} - \hat{x}_{IIB}(0)\big|_{\beta=1} = \frac{n(n-1)^2 b\gamma\theta^2(a-A)}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - 2n^2(1-\theta)[1+\theta(n-2)]}$$

$$\times \frac{[1+\theta(n-1)]}{2b\gamma[1+\theta(n-1)]-n^2} > 0.$$

From the proof of part (i) it then readily follows that also $\forall n, \beta, \theta$,

$$\hat{x}_{III}(0)\big|_{\beta=1} > \hat{x}_{IC}(0), \, \hat{x}_{IIB}(0), \, \hat{x}_{IIC}(0).$$

However, as shown above, under second-stage Bertrand competition, competitive R&D efforts do not always fall below those of an RJV-cartel. To be considered in addition therefore is

$$\hat{x}_{III}(0)\big|_{\beta=1} - \hat{x}_{IB}(0) = \frac{(a-A)}{2b\gamma[1+\theta(n-1)]-n^2}$$

$$\times \frac{b\gamma[1+\theta(n-1)]\{4(n-1)\theta\beta[1+\theta(n-2)]^2 + 2\Delta n(n-1)(1-\beta)[1+\theta(n-2)] + \alpha_{71}^B + \alpha_{72}^B\}}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 - 2\Delta[1+\theta(n-2)][1+\beta(n-1)]}.$$

with $\alpha_{71}^B = [2+\theta(2n-3)][2+\theta(n-3)]^2 - 4[1+\theta(n-2)]\{2+3\theta(n-2)+\theta^2[(n-1)(n-2)-(2n-3)]\}$, and $\alpha_{72}^B = (n-1)[2+\theta(2n-3)][2+\theta(n-3)]^2$. A sufficient condition for $\hat{x}_{III}(0)\big|_{\beta=1} - \hat{x}_{IB}(0)$ to be positive is that $\alpha_{71}^B + \alpha_{72}^B$ is greater than zero. Then observe that $\forall n, \theta, \beta$,

$$\alpha_{71}^B + \alpha_{72}^B > 2[1+\theta(n-2)]\{4(n-1)+2\theta[2n^2-9n+6]+\theta^2[n^3-8n^2+19n-10]\} > 0.$$

The proof of the second part of the lemma is complete if it is realized that $\forall n, \beta, \theta$,

$$\frac{\partial \hat{x}_{III}(0)}{\partial \beta} = \frac{(n-1)(a-A)\{2b\gamma[1+\theta(n-1)]+[1+\beta(n-1)]^2\}}{\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}^2} > 0.$$

Q.E.D.

2B.3 Proof Lemma 2.3

Comparing the levels of non-subsidized profits leads to

$$\hat{\pi}_{III}(0) - \hat{\pi}_{IIC}(0) = \frac{\gamma(n-1)^2(a-A)^2 b\gamma\theta^2}{2\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}\{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2\}},$$

$$\hat{\pi}_{IIC}(0) - \hat{\pi}_{IC}(0) = \frac{2n\gamma(a-A)^2 b\gamma}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2}$$

$$\times \frac{(n-1)^2[2+\theta(n-1)]^2[2\beta-\theta]^2}{\{b\gamma(2-\theta)[2+\theta(n-1)]^2-2\{[2+\theta(n-2)]-\theta\beta(n-1)\}[1+\beta(n-1)]\}^2},$$

and

$$\hat{R}_{III}(0) - \hat{R}_{IIB}(0) = \frac{\gamma(n-1)^2(a-A)^2}{2\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}}$$

$$\times \frac{b\gamma\theta^2[1+\theta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2},$$

$$\hat{R}_{IIB}(0) - \hat{R}_{IB}(0) = \frac{2\gamma(n-1)^2(a-A)^2}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2}$$

$$\times \frac{b\gamma[1+\theta(n-1)][1+\theta(n-2)]^2[2+\theta(n-3)]^2\{\beta[2+2\theta(n-2)-\theta^2(n-1)]-\theta[1+\theta(n-2)]\}^2}{\{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]\}^2},$$

from which the lemma follows immediately.

<div align="right">*Q.E.D.*</div>

2B.4 Proof Lemma 2.4

The appropriate partial derivatives equal

$$\frac{\partial \hat{R}_{IIC}(0)}{\partial \beta} = \frac{4\gamma(n-1)(a-A)^2[1+\beta(n-1)]}{\{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2\}^2} > 0,$$

$$\frac{\partial \hat{R}_{IIB}(0)}{\partial \beta} = \frac{4\gamma(n-1)(a-A)^2(1-\theta)^2[1+\theta(n-2)]^2[1+\beta(n-1)]}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2\}^2} > 0,$$

$$\frac{\partial \hat{R}_{III}(0)}{\partial \beta} = \frac{4\gamma(n-1)(a-A)^2[1+\beta(n-1)]}{4\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}^2} > 0.$$

<div align="right">*Q.E.D.*</div>

2B.5 Proof Lemma 2.5

First observe that net total surplus conditional on the R&D investment is the same for the first and second scenario considered. As shown in Lemma 2.7, under second-stage Bertrand competition optimal R&D-subsidies are always positive. Hence, both $\hat{x}_{IB}(0)$ and $\hat{x}_{IIB}(0)$ are below x_B^*, the latter being the R&D activity maximizing net total surplus under either the first or the second regime. The lemma then follows immediately.

If, under second-stage Cournot competition, the optimal R&D-subsidies are positive, the same reasoning applies. However, as shown in Lemma 2.7, there are circumstances in which the optimal R&D-subsidy for the competitive game is negative. This means that $\hat{x}_{IC}(0)$ could exceed x_C^*. Depending on specific parameter configurations, either net total surplus under full competition exceeds that under partial collusion, or the latter exceeds the former.

<div align="right">*Q.E.D.*</div>

2B.6 Proof Lemma 2.6

The appropriate partial derivatives equal

$$\frac{\partial \hat{W}_{IIC}(0)}{\partial \beta} = \frac{4\gamma n(n-1)(a-A)^2[1+\beta(n-1)]\{b\gamma[2+\theta(n-1)]^3 - 2[1+\beta(n-1)]^2\}}{\{b\gamma[2+\theta(n-1)]^2 - 2[1+\beta(n-1)]^2\}^3} > 0,$$

$$\frac{\partial \hat{W}_{IIB}(0)}{\partial \beta} = 4\gamma n(n-1)(a-A)^2(1-\theta)[1+\theta(n-2)]^2[1+\beta(n-1)]$$

$$\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^3 - 2(1-\theta)^2[1+\theta(n-2)][1+\beta(n-1)]^2}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - 2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2\}^3} > 0,$$

$$\frac{\partial \hat{W}_{III}(0)}{\partial \beta} = \frac{\gamma n(n-1)(a-A)^2[1+\beta(n-1)]\{4b\gamma[1+\theta(n-1)] - [1+\beta(n-1)]^2\}}{\{2b\gamma[1+\theta(n-1)] - [1+\beta(n-1)]^2\}^3} > 0.$$

Q.E.D.

2B.7 Proof Lemma 2.7

Part (i)

This follows immediately from the second-order conditions associated with deriving the optimal R&D-subsidies.

Part (ii)

For the optimal R&D-subsidy to be positive it must be that

$$(2-\theta)[3+\theta(n-1)][1+\beta(n-1)] - 2[2+\theta\{(n-2)-\beta(n-1)\}] > 0,$$

or

$$\beta > \frac{-2+\theta+\theta^2(n-1)}{(n-1)\{6+\theta(2n-3)-\theta^2(n-1)\}}. \tag{2B.1}$$

Differentiating the RHS of (2B.1) with respect to θ leads to

$$\frac{4n+8\theta(n-1)+2\theta^2(n-1)^2}{(n-1)\{6+\theta(2n-3)-\theta^2(n-1)\}^2},$$

which is obviously positive. Hence a *sufficient* condition for s_{IC}^{*} to be positive is

$$\beta > \frac{(n-2)}{(n-1)(n+4)}. \tag{2B.2}$$

The RHS of (2B.2) reaches its maximum for $n=4$ or $n=5$. Hence, β should exceed $1/12$.

Q.E.D.

2B.8 Proof Lemma 2.8

The appropriate partial derivatives equal

$$\frac{\partial s_{IC}^{\bullet}}{\partial \beta} = \frac{\gamma(n-1)(a-A)\{[6+\theta(2n-3)-\theta^2(n-1)]\alpha_{71}^{C}+2[3+\theta(n-1)]\alpha_{72}^{C}\}}{(2-\theta)\left[\alpha_{71}^{C}\right]^2} > 0,$$

$$\frac{\partial s_{IIC}^{\bullet}}{\partial \beta} = \frac{\gamma(n-1)(a-A)[1\theta(n-1)]\{b\gamma[2+\theta(n-1)]^2+[3+\theta(n-1)][1+\beta(n-1)]^2\}}{\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2} > 0,$$

with $\alpha_{71}^{C} = b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2$, and
$\alpha_{72}^{C} = (2-\theta)[3+\theta(n-1)][1+\beta(n-1)]-2[2+\theta\{(n-2)-\beta(n-1)\}]$, and

$$\frac{\partial s_{IB}^{\bullet}}{\partial \beta} = \frac{\gamma(n-1)(a-A)\{[6+\theta(8n-15)+\theta^2(2n^2-9n+8)\alpha_{71}^{B}+2[1+\theta(n-2)][3+\theta(n-1)]\alpha_{72}^{B}\}}{[2+\theta(2n-3)]\left[\alpha_{72}^{B}\right]^2} > 0,$$

$$\frac{\partial s_{IIB}^{\bullet}}{\partial \beta} = \gamma(n-1)(a-A)[1+\theta(n-2)]^2$$

$$\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2+[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2} > 0,$$

with $\alpha_{71}^{B} = b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-\{1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2$, and
$\alpha_{72}^{B} = [2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]-2\Delta$, and

$$\frac{\partial s_{III}^{\bullet}}{\partial \beta} = \frac{\gamma(n-1)(a-A)\{4b\gamma[1+\theta(n-1)]+3[1+\beta(n-1)]^2\}}{\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}^2} > 0.$$

Q.E.D.

2B.9 Proof of Lemma 2.9

Comparing the different R&D-subsidies under the first two regimes leads to

$$s_{IC}^{\bullet}-s_{IIC}^{\bullet} = \frac{2\gamma(n-1)(a-A)[2\beta-\theta]}{(2-\theta)\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}},$$

$$s_{IB}^{\bullet}-s_{IIB}^{\bullet} = \frac{2\gamma(n-1)(a-A)[1+\theta(n-2)]}{[2+\theta(2n-3)]}$$

$$\times \frac{\beta[2+2\theta(n-2)-\theta^2(n-1)]-\theta[1+\theta(n-2)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}.$$

Q.E.D.

2B.10 Proof of Lemma 2.10

Part (i)

$\forall n, \beta, \theta$, the following holds

$$\overset{\bullet}{x}_{IC} - \overset{\bullet}{x}_{III} = \frac{b\gamma(n-1)(a-A)}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}$$

$$\times \frac{\theta[4+\theta(n-1)][1+\beta(n-1)]}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2} > 0.$$

$$\overset{\bullet}{x}_{IB} - \overset{\bullet}{x}_{III} = \frac{b\gamma(n-1)(a-A)}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$\times \frac{\theta[1+\theta(n-1)][4+\theta(n-5)][1+\beta(n-1)]}{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2} > 0.$$

Part (ii)

The appropriate partial derivatives equal

$$\frac{\partial \overset{\bullet}{x}_{IC}}{\partial \beta} = (n-1)(a-A)[3+\theta(n-1)]$$

$$\times \frac{b\gamma[2+\theta(n-1)]^2+[3+\theta(n-1)][1+\beta(n-1)]^2}{\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2} > 0.$$

$$\frac{\partial \overset{\bullet}{x}_{IB}}{\partial \beta} = (n-1)(a-A)[1+\theta(n-2)][3+\theta(n-4)]$$

$$\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2+[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2} > 0.$$

Q.E.D.

2B.11   Proof of Lemma 2.11

Part (i)

Simulation results reveal the following

$$\min_{n,\beta,\theta}\{\pi_{III}^{*}-\pi_{IC}^{*}\} > 10^{-3},$$

$$\min_{n,\beta,\theta}\{\pi_{III}^{*}-\pi_{IB}^{*}\} > 10^{-3}.$$

Part (ii)

The appropriate partial derivatives equal

$$\frac{\partial \pi_{IC}^{*}}{\partial \beta} = -\gamma(n-1)b\gamma(a-A)^{2}$$

$$\times \frac{[1+\theta(n-1)][3+\theta(n-1)][2+\theta(n-1)]^{2}[1+\beta(n-1)]}{\{b\gamma[2+\theta(n-1)]^{2}-[3+\theta(n-1)][1+\beta(n-1)]^{2}\}^{3}} < 0,$$

$$\frac{\partial \pi_{IB}^{*}}{\partial \beta} = -\gamma(n-1)b\gamma(a-A)^{2}$$

$$\times \frac{[1+\theta(n-1)][3+\theta(n-4)][2+\theta(n-3)]^{2}[1+\theta(n-2)]^{3}[1+\beta(n-1)]}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^{2}-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^{2}\}^{3}} < 0,$$

$$\frac{\partial \pi_{III}^{*}}{\partial \beta} = \frac{3\gamma(n-1)(a-A)^{2}[1+\beta(n-1)]\{4b\gamma[1+\theta(n-1)]-9[1+\beta(n-1)]^{2}\}}{\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^{2}\}^{3}} > 0.$$

Q.E.D.

2B.12 Proof of Lemma 2.12

Part (i)

$\forall n, \beta, \theta$, the following holds

$$W_{IC}^* - W_{III}^* = \frac{\gamma n(n-1)b\gamma(a-A)^2}{2\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}}$$

$$\times \frac{\theta[4+\theta(n-1)]}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2} > 0.$$

$$W_{IB}^* - W_{III}^* = \frac{\gamma n(n-1)b\gamma(a-A)^2}{2\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}}$$

$$\times \frac{\theta[1+\theta(n-1)][4+\theta(n-5)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}} > 0.$$

Part (ii)

The appropriate partial derivatives equal

$$\frac{\partial W_{IC}^*}{\partial \beta} = \frac{\gamma n(n-1)(a-A)^2[3+\theta(n-1)]^2}{\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2} > 0,$$

$$\frac{\partial W_{IB}^*}{\partial \beta} = \frac{\gamma n(n-1)(a-A)^2[1+\theta(n-2)]^2[3+\theta(n-4)]^2}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2} > 0,$$

$$\frac{\partial W_{III}^*}{\partial \beta} = \frac{9\gamma n(n-1)(a-A)^2[1+\beta(n-1)]}{\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}^2} > 0.$$

Q.E.D.

Appendix 2C Proofs of propositions

2C.1 Proof of Proposition 2.1

R&D

Comparing subsidized and non-subsidized R&D under second-stage Cournot and Bertrand competition leads to

$$
x_{IC}^{*} - \hat{x}_{IC}(0) = \frac{b\gamma(a-A)[2+\theta(n-1)]^2}{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2}
$$

$$
\times \frac{(2-\theta)[1+\theta\beta(n-1)^2]+(n-1)[6\beta-\theta(\theta+\beta)]}{b\gamma(2-\theta)[2+\theta(n-1)]^2-2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]} > 0,
$$

$$
x_{IIC}^{*} - \hat{x}_{IIC}(0) = \frac{b\gamma(a-A)[1+\theta(n-1)][2+\theta(n-1)]^2[1+\beta(n-1)]}{\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}\{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2\}} > 0,
$$

$$
x_{IB}^{*} - \hat{x}_{IB}(0) = \frac{b\gamma(a-A)[1+\theta(n-2)][1+\theta(n-1)][2+\theta(n-3)]^2}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2-2\Delta[1+\theta(n-2)][1+\beta(n-1)]}
$$

$$
\times \frac{(2-\theta)[1+\theta(n-2)]+\beta(n-1)[6+\theta(8n-15)+\theta^2[2n^2-9n+8]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2} > 0,
$$

$$
x_{IIB}^{*} - \hat{x}_{IIB}(0) = \frac{b\gamma(a-A)[1+\theta(n-2)]^2[1+\theta(n-1)][2+\theta(n-3)]^2[1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}
$$

$$
\times \frac{1}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2} > 0,
$$

$$
x_{III}^{*} - \hat{x}_{III}(0) = \frac{2b\gamma(a-A)[1+\theta(n-1)][1+\beta(n-1)]}{\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}} > 0.
$$

The first of these inequalities hold if $n=2$, or $n\geq3 \wedge \beta>1/12$, as shown in the proof of Lemma 2.7. All other inequalities hold $\forall n,\beta,\theta$.

Output

Comparing subsidized and non-subsidized output under second-stage Cournot and Bertrand competition leads to

$$Q_{IC}^{*} - \hat{Q}_{IC}(0) = \frac{n\gamma(a-A)[2+\theta(n-1)][1+\beta(n-1)]}{b\gamma[2+\theta(n-1)]^2 - [3+\theta(n-1)][1+\beta(n-1)]^2}$$

$$\times \frac{(2-\theta)[3+\theta(n-1)][1+\beta(n-1)] - 2[2+\theta\{(n-2)-\beta(n-1)\}]}{b\gamma(2-\theta)[2+\theta(n-1)]^2 - 2[1+\beta(n-1)][2+\theta\{(n-2)-\beta(n-1)\}]} > 0.$$

$$Q_{IIC}^{*} - \hat{Q}_{IIC}(0) = \frac{n\gamma(a-A)[1+\theta(n-1)][2+\theta(n-1)][1+\beta(n-1)]^2}{\{b\gamma[2+\theta(n-1)]^2 - [3+\theta(n-1)][1+\beta(n-1)]^2\}\{b\gamma[2+\theta(n-1)]^2 - 2[1+\beta(n-1)]^2\}} > 0.$$

$$Q_{IB}^{*} - \hat{Q}_{IB}(0) = \frac{n\gamma(a-A)[2+\theta(n-3)][1+\theta(n-2)]^2[1+\beta(n-1)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - [1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$\times \frac{(2-\theta)[1+\theta(n-2)] + \beta(n-1)[6+\theta(8n-15)+\theta^2(2n^2-9n+8)]}{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 - 2\Delta[1+\theta(n-2)][1+\beta(n-1)]} > 0.$$

$$Q_{IIB}^{*} - \hat{Q}_{IIB}(0) = \frac{n\gamma(a-A)[2+\theta(n-3)][1+\theta(n-2)]^3[1+\beta(n-1)]^2}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - [1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}$$

$$\times \frac{1}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - 2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2} > 0,$$

$$Q_{III}^{*} - \hat{Q}_{III}(0) = \frac{n\gamma(a-A)[1+\beta(n-1)]^2}{\{4b\gamma[1+\theta(n-1)] - 3[1+\beta(n-1)]^2\}\{2b\gamma[1+\theta(n-1)] - [1+\beta(n-1)]^2\}} > 0.$$

The first of these inequalities hold if $n=2$, or $n \geq 3 \wedge \beta > 1/12$, as shown in the proof of Lemma 2.7. All other inequalities hold $\forall n, \beta, \theta$.

Prices

The fact that prices under the subsidized regimes are lower than under the concomitant non-subsidized regimes follows readily from (2.2) combined with the effects the provision of optimal R&D-subsidies (with a concomitant tax in the production stage) have on output.

Net Total Surplus

Comparing subsidized and non-subsidized net total surplus under second-stage Cournot and Bertrand competition leads to

$$W_{IC}^* - \hat{W}_{IC}(0) = \frac{n\gamma(a-A)^2 b\gamma[2+\theta(n-1)]^2}{2\{b\gamma[2+\theta(n-1)]^2 - [3+\theta(n-1)][1+\beta(n-1)]^2\}}$$

$$\times \frac{\{(2-\theta)[3+\theta(n-1)][1+\beta(n-1)] - 2[2+\theta\{(n-2)-\beta(n-1)\}]\}^2}{\{b\gamma(2-\theta)[2+\theta(n-1)]^2 - 2[2+\theta\{(n-2)-\beta(n-1)\}][1+\beta(n-1)]\}^2} > 0.$$

$$W_{IIC}^* - \hat{W}_{IIC}(0) = \frac{n\gamma(a-A)^2}{2\{b\gamma[2+\theta(n-1)]^2 - [3+\theta(n-1)][1+\beta(n-1)]^2\}}$$

$$\times \frac{b\gamma[1+\theta(n-1)]^2[2+\theta(n-1)]^2[1+\beta(n-1)]^2}{\{b\gamma[2+\theta(n-1)]^2 - 2[1+\beta(n-1)]^2\}^2} > 0.$$

$$W_{IB}^* - \hat{W}_{IB}(0) = \frac{n\gamma(a-A)^2 b\gamma[1+\theta(n-1)][1+\theta(n-2)]^2[2+\theta(n-3)]^2}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - [1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}}$$

$$\times \frac{\{2\Delta - [2+\theta(2n-3)][3+\theta(n-4)][1+\beta(n-1)]\}^2}{\{b\gamma[1+\theta(n-1)][2+\theta(2n-3)][2+\theta(n-3)]^2 - 2\Delta[1+\theta(n-2)][1+\beta(n-1)]^2\}^2} > 0.$$

$$W_{IIB}^* - \hat{W}_{IIB}(0) = \frac{n\gamma(a-A)^2}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - [1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}}$$

$$\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2[1+\theta(n-2)]^4[1+\beta(n-1)]^2}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2 - 2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2\}^2} > 0.$$

$$W_{III}^* - \hat{W}_{III}(0) = \frac{n\gamma(a-A)^2 b\gamma[1+\theta(n-1)][1+\beta(n-1)]^2}{2\{4b\gamma[1+\theta(n-1)] - 3[1+\beta(n-1)]^2\}\{2b\gamma[1+\theta(n-1)] - [1+\beta(n-1)]^2\}^2} > 0.$$

The first of these inequalities hold if $n=2$, or $n\geq3 \wedge \beta>1/12$, as shown in the proof of Lemma 2.7. All other inequalities hold $\forall n, \beta, \theta$.

Profits

Comparing subsidized and non-subsidized profits under second-stage Cournot and Bertrand competition leads to

$$\pi_{IIC}^* - \hat{\pi}_{IIC}(0) = - \frac{\gamma(a-A)^2}{2\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2}$$

$$\times \frac{b\gamma[1+\theta(n-1)]^2[2+\theta(n-1)]^2[1+\beta(n-1)]^2}{b\gamma[2+\theta(n-1)]^2-2[1+\beta(n-1)]^2} < 0,$$

$$\pi_{IIB}^* - \hat{\pi}_{IIC}(0) = - \frac{\gamma(a-A)^2}{2\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2}$$

$$\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2[1+\theta(n-2)]^4[1+\beta(n-1)]^2}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(1-\theta)[1+\theta(n-2)][1+\beta(n-1)]^2} < 0,$$

$$\pi_{III}^* - \hat{\pi}_{III}(0) = - \frac{\gamma(a-A)^2 b\gamma[1+\theta(n-1)][1+\beta(n-1)]^2}{\{4b\gamma[1+\theta(n-1)]-3[1+\beta(n-1)]^2\}^2\{2b\gamma[1+\theta(n-1)]-[1+\beta(n-1)]^2\}} < 0,$$

which all hold $\forall n, \beta, \theta$. For the two noncooperative cases numerical simulations have to be relied upon. These show that

$$\min_{n>3, \beta \in (1/12,], \theta} \{\hat{\pi}_{IC}(0) - \pi_{IC}^*\} > 10^{-5},$$

$$\min_{n, \beta, \theta} \{\hat{\pi}_{IB}(0) - \pi_{IB}^*\} > 10^{-7}.$$

Q.E.D.

2C.2 Proof of Proposition 2.2

Part (i)

Comparing the equilibrium levels of noncooperative, subsidized R&D with the non-subsidized R&D investment in an R&D-cartel reveals that

$$
x_{IC}^* = \hat{x}_{IIC}(0) \left\{ 1 + \frac{b\gamma[1+\theta(n-1)][2+\theta(n-1)]^2}{2\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}} \right\},
$$

$$
x_{IB}^* = \hat{x}_{IIB}(0) \left\{ 1 + \frac{b\gamma[1+\theta(n-1)][1+\theta(n-2)][2+\theta(n-3)]^2}{2(1-\theta)\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}} \right\}.
$$

Appropriate second-order conditions ensure that the expressions in brackets exceed 1.

Part (ii)

The partial derivatives of noncooperative, subsidized R&D with respect to β equal

$$
\frac{\partial x_{IC}^*}{\partial \beta} = \frac{(a-A)(n-1)[3+\theta(n-1)]\{b\gamma[2+\theta(n-1)]^2+[3+\theta(n-1)][1+\beta(n-1)]^2\}}{\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)][1+\beta(n-1)]^2\}^2} > 0,
$$

$$
\frac{\partial x_{IB}^*}{\partial \beta} = (a-A)(n-1)[1+\theta(n-2)][3+\theta(n-4)]
$$

$$
\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2+[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2}{\{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)][1+\beta(n-1)]^2\}^2} > 0.
$$

The proof is complete if it is realized that

$$
x_{IC}^*|_{\beta=0} - \hat{x}_{IC}(0)|_{\beta=1} = \frac{(a-A)\{b\gamma[1+\theta(n-1)][2+\theta(n-1)]^2-2(n-1)[3+\theta(n-1)]\}}{\{b\gamma[2+\theta(n-1)]^2-[3+\theta(n-1)]\}\{b\gamma[2+\theta(n-1)]^2-2n\}} > 0,
$$

$$
x_{IB}^*|_{\beta=0} - \hat{x}_{IB}(0)|_{\beta=1} = \frac{(a-A)[1+\theta(n-2)]^2}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-[1+\theta(n-2)][3+\theta(n-4)]}
$$

$$
\times \frac{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2(n-1)(1-\theta)[3+\theta(n-4)]}{b\gamma[1+\theta(n-1)][2+\theta(n-3)]^2-2n(1-\theta)[1+\theta(n-2)]} > 0.
$$

Q.E.D.

Appendix 2D Simulations

2D.1 Description of the numerical procedures

For any numerical procedure, the values of a,A and b are fixed. In particular, $(a-A)=10$ and $b=1$. Given these, γ is set such that all second-order conditions, stability conditions and the conditions for the tax rates to be below 1, hold.

The numerical approximations are either of two types. The first are numerical proofs of statements which are analytically not tractable. In these cases, the complete parameter space of θ (being $(0,1)$) and β (being $[0,1]$) is considered for $2 \leq n \leq 50$. The stepsize for θ and β is 0.01. The upper bound on the size of the industry considered is only chosen as to keep the computations within reasonable time limits. Enlarging n does not change qualitatively any of the results reported.

The second type of numerical simulations consider those expressions which do not have a unique sign. For these situations an overview is presented considering the sign of the expression of interest for different values of β, θ and n. It is these overviews which are contained in the remainder of this appendix.

2D.2 Simulation results

Table 2D.1 Simulated signs; $\partial \hat{\pi}_i(0)/\partial \beta$

n	θ (β)	Cournot competition					Bertrand competition				
		0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
	0.01	+	+	+	+	+	+	+	+	+	+
	0.25	+	+	+	+	+	+	+	+	+	+
2	0.50	+	+	+	+	+	+	+	+	+	+
	0.75	+	+	+	+	+	+	+	+	+	-
	0.99	+	+	+	+	+	+	+	+	+	-
	0.01	+	+	+	+	+	+	+	+	+	+
	0.25	+	+	+	+	+	+	+	+	+	+
3	0.50	+	+	+	+	+	+	+	+	+	+
	0.75	+	+	+	+	-	+	+	+	+	-
	0.99	+	+	+	+	-	+	+	+	+	-
	0.01	+	+	+	+	+	+	+	+	+˙	+
	0.25	+	+	+	+	-	+	+	+	+	-
8	0.50	+	+	+	+	-	+	+	+	-	-
	0.75	+	+	+	-	-	+	+	+	-	-
	0.99	+	+	+	-	-	+	+	+	+	-
	0.01	+	+	+	+	+	+	+	+	+	+
	0.25	+	+	+	-	-	+	+	+	-	-
20	0.50	+	+	+	-	-	+	+	+	-	-
	0.75	+	+	+	-	-	+	+	+	-	-
	0.99	+	+	+	-	-	+	+	+	+	-
	0.01	+	+	+	+	+	+	+	+	+	+
	0.25	+	+	+	-	-	+	+	+	-	-
50	0.50	+	+	+	-	-	+	+	+	-	-
	0.75	+	+	+	-	-	+	+	+	-	-
	0.99	+	+	+	-	-	+	+	+	+	-

Table 2D.2 Simulated signs; $\partial \hat{W}_i(0)/\partial \beta$

n	θ	β	0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
	Second stage		Cournot competition					Bertrand competition				
	0.01		+	+	+	+	+	+	+	+	+	+
	0.25		+	+	+	+	+	+	+	+	+	+
2	0.50		+	+	+	+	+	+	+	+	+	+
	0.75		+	+	+	+	+	+	+	+	-	-
	0.99		+	+	+	+	-	+	+	-	-	-
	0.01		+	+	+	+	+	+	+	+	+	+
	0.25		+	+	+	+	+	+	+	+	+	+
3	0.50		+	+	+	+	+	+	+	+	+	-
	0.75		+	+	+	+	-	+	+	+	-	-
	0.99		+	+	+	-	-	+	+	-	-	-
	0.01		+	+	+	+	+	+	+	+	+	+
	0.25		+	+	+	+	+	+	+	+	+	-
8	0.50		+	+	+	-	-	+	+	+	-	-
	0.75		+	+	+	-	-	+	+	+	-	-
	0.99		+	+	+	-	-	+	+	-	-	-
	0.01		+	+	+	+	+	+	+	+	+	+
	0.25		+	+	+	-	-	+	+	+	-	-
20	0.50		+	+	+	-	-	+	+	+	-	-
	0.75		+	+	+	-	-	+	+	+	-	-
	0.99		+	+	+	-	-	+	+	-	-	-
	0.01		+	+	+	+	+	+	+	+	+	+
	0.25		+	+	+	-	-	+	+	+	-	-
50	0.50		+	+	+	-	-	+	+	+	-	-
	0.75		+	+	+	-	-	+	+	+	-	-
	0.99		+	+	+	-	-	+	+	-	-	-

Table 2D.3 Simulated signs; $\hat{W}_I(0)-\hat{W}_{III}(0)|_{\beta=1}$ [a]

n	θ	β	Cournot competition					Bertrand competition				
			0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
	0.01		-	-	-	-	-	-	-	-	-	+
	0.25		+	+	+	+	+	+	+	+	+	+
2	0.50		+	+	+	+	+	+	+	+	+	+
	0.75		+	+	+	+	+	+	+	+	+	+
	0.99		+	+	+	+	+	+	+	+	+	+
	0.01		-	-	-	-	-	-	-	-	-	+
	0.25		+	+	+	+	+	+	+	+	+	+
3	0.50		+	+	+	+	+	+	+	+	+	+
	0.75		+	+	+	+	+	+	+	+	+	+
	0.99		+	+	+	+	+	+	+	+	+	+
	0.01		-	-	-	+	+	-	-	-	+	+
	0.25		-	+	+	+	+	-	+	+	+	+
8	0.50		+	+	+	+	+	+	+	+	+	+
	0.75		+	+	+	+	+	+	+	+	+	+
	0.99		+	+	+	+	+	+	+	+	+	+
	0.01		+	-	+	+	+	+	-	+	+	+
	0.25		-	+	+	+	+	-	+	+	+	+
20	0.50		-	+	+	+	+	-	+	+	+	+
	0.75		-	+	+	+	+	-	+	+	+	+
	0.99		-	+	+	+	+	-	+	+	+	+
	0.01		+	-	+	+	+	+	-	+	+	+
	0.25		+	+	+	+	+	+	+	+	+	+
50	0.50		-	+	+	+	+	-	+	+	+	+
	0.75		-	+	+	+	+	-	+	+	+	+
	0.99		-	+	+	+	+	-	+	+	+	+

[a] For $n \geq 129$ all entries are positive.

Table 2D.4 Simulated signs; $s_I^{\bullet} - s_{III}^{\bullet}$

n	θ	β		Cournot competition					Bertrand competition			
			0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
	0.01		+	+	+	+	+	+	+	+	+	+
	0.25		-	+	+	+	+	+	+	+	+	+
2	0.50		-	+	+	+	+	+	+	+	+	+
	0.75		-	+	+	+	+	+	+	+	+	+
	0.99		-	+	+	+	+	+	+	+	+	+
	0.01		-	+	+	+	+	+	+	+	+	+
	0.25		-	+	+	+	+	-	+	+	+	+
3	0.50		-	+	+	+	+	-	+	+	+	+
	0.75		-	+	+	+	+	-	+	+	+	+
	0.99		-	+	+	+	+	-	+	+	+	+
	0.01		-	+	+	+	+	-	+	+	+	+
	0.25		-	+	+	+	+	-	+	+	+	+
8	0.50		-	+	+	+	+	-	+	+	+	+
	0.75		-	+	+	+	+	-	+	+	+	+
	0.99		-	+	+	+	+	-	+	+	+	+
	0.01		-	+	+	+	+	-	+	+	+	+
	0.25		-	+	+	+	+	-	+	+	+	+
20	0.50		-	+	+	+	+	-	+	+	+	+
	0.75		-	+	+	+	+	-	+	+	+	+
	0.99		-	+	+	+	+	-	+	+	+	+
	0.01		-	+	+	+	+	-	+	+	+	+
	0.25		-	+	+	+	+	-	+	+	+	+
50	0.50		-	+	+	+	+	-	+	+	÷	+
	0.75		-	+	+	+	+	-	+	+	+	+
	0.99		-	+	+	+	+	-	+	+	+	+

3 On the equivalence of subsidized cooperative and noncooperative R&D

3.1 Introduction

The previous chapter compared two widely used policies to stimulate private R&D: allowing firms to cooperate in their research, and the provision of direct R&D-subsidies (for which firms are taxed in the product market). Moreover, the implementation of both policies simultaneously was considered. The analysis lead to some remarkable results. Not only is the provision of direct R&D-subsidies in most cases more effective in raising private R&D spending than allowing firms to cooperate in R&D, but subsidizing cooperative or noncooperative R&D leads to exactly the same market outcome. Hence, to promote private R&D efforts, allowing firms to cooperate in the pre-competitive stage of the production process is redundant.

Recall however that the analysis of Chapter 2 was carried out within a linear-quadratic framework. And although specific functional forms of demand and cost functions allow many detailed results, the question remains as to what extent conclusions are robust with respect to the choice of particular functional forms. This chapter addresses the same issues as those raised in Chapter 2 within a general setting. In particular, the framework developed by Suzumura [1992] is built upon.[1] Within an oligopolistic framework with general demand and cost functions he considers the issue of cooperative R&D when firms are engaged in second-stage Cournot competition. In addition, we allow products to be differentiated, so that second-stage Bertrand competition can also be analyzed, and consider a more general R&D cost function than the one used by Suzumura [1992].[2]

Given the limited scope of this study not all results of Chapter 2 will be generalized. In particular, the case in which firms collude in the product market will not be considered. Rather, the focus will be on comparing the two R&D stimulating policies, and on considering the effects of implementing both policies simultaneously. The main result is an equivalence theorem, which states that subsidizing cooperative or noncooperative R&D optimally (under the provision that firms are taxed for the R&D-subsidy in the product market) leads to the same market outcome and social welfare. Indeed, the equivalence theorem is a general version of Proposition 2.3. Making then a mild assumption as to the cost of R&D and assuming products to be homogeneous, a corollary of the equivalence theorem is derived under second-stage Cournot competition: subsidizing competitive R&D optimally is more effective in raising private R&D efforts than allowing firms to form an R&D-cartel, which is a general version of part of Proposition 2.2.

---

[1] His model is also used by Simpson and Vonortas [1994].

[2] The focus of the analysis will be on subsidizing R&D versus allowing for R&D-cooperatives, since obtaining comparative statics results of multi-stage games with Bertrand competition, given that products are differentiated, within a general oligopolistic framework might well be impossible (see Dixit [1984]).

In what follows, first the analytical framework is described. The noncooperative and cooperative market equilibria are characterized in Sections 3.3 and 3.4. The equivalence theorem is then derived, followed by a comparison of the two R&D stimulating policies. Section 3.7 concludes.

3.2 The model

3.2.1 Demand

The representative consumer is assumed to have preferences over the consumption bundle $q = \{q_1,...,q_n\}$, which are complete, reflexive, transitive, continuous, and strongly monotonic. Hence, there exists a continuous utility function $U(\cdot): \mathbf{R}_+^n \to \mathbf{R}_+$, which represents these preferences (see e.g. Varian [1984, p.113]). In addition it is assumed that preferences are strictly convex, or that the utility function is strictly concave. Also, each commodity $q_i \in q$ is assumed to be differentiated from all other goods $q_j \in q$, $j \neq i$. The representative consumer is assumed to maximize $U(\cdot)$ over q, given that the upper limit to it's budget is given by $\sum_{i=1}^{n} p_i q_i$, where p_i is the price of commodity i. Under strongly monotonic preferences and assuming that money does not enter $U(\cdot)$, there are no savings, that is, all disposable income is spent.

The representative consumer's maximizing behaviour leads to a system of inverse demands given by

$$p(q) = \{p_1(q),...,p_n(q)\}. \tag{3.1}$$

Under the assumptions on the representative consumer's preference, $p(q)$ is twice continuously differentiable with $\partial p_i(q)/\partial q < 0$ for all q satisfying $p_i(q) > 0$. Moreover, given that products are differentiated, (3.1) is invertible. Hence, direct demand,

$$q(p) = \{q_1(p),...,q_n(p)\}, \tag{3.2}$$

is well defined and also twice continuously differentiable.

3.2.2 Supply

As in Chapter 2, the industry under consideration comprises n firms, each of which produces one variety of the differentiated commodity. Fixed costs of production are assumed to be the same for all firms and are set equal to zero. Firms can devote resources to process-innovating R&D, that is, firm i's marginal costs of production, $c_i(\cdot): \mathbf{R}_+^n \to \mathbf{R}_+$ can be reduced by conducting research. Hence,

$$\partial c_i(x)/\partial x_i < 0, \tag{3.3}$$

where $x = \{x_1,...,x_n\}$ is the vector of R&D activity in the industry considered. It is assumed that $c_i(x)$ is twice continuously differentiable. Further, there are technological spillovers, which means that
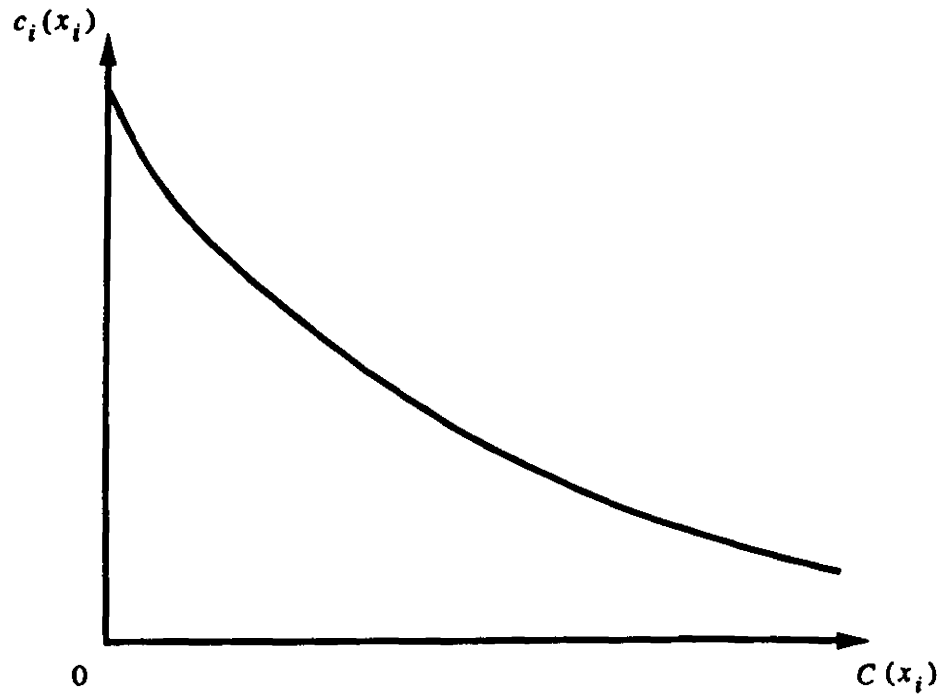
Figure 3.1 Diminishing returns to scale in R&D

$$\partial c_i(x)/\partial x_j < 0. \tag{3.4}$$

Following Suzumura [1992], it is assumed that

$$|\partial c_i(x)/\partial x_i| \geq |\partial c_i(x)/\partial x_j|. \tag{3.5}$$

In terms of Chapter 2 this means that the spillover rate β will be less than or equal to 1. Indeed, in an RJV or RJV-cartel we have that $|\partial c_i(x)/\partial x_i| = |\partial c_i(x)/\partial x_j|$, for $i \neq j$.

The cost of investing in R&D, $C_i(\cdot): \mathbf{R}_+ \to \mathbf{R}_+$, is assumed to be the same for all firms. Hence, $C_i(\cdot) = C_j(\cdot) = C(\cdot)$. This R&D cost function is also assumed to be twice continuously differentiable, with, in particular

$$\partial C(x_i)/\partial x_i > 0. \tag{3.6}$$

Observe that the assumption as to whether or not the process of R&D exhibits decreasing returns to scale is implicit in the relation between the marginal cost of production function and the R&D cost function.[3] In particular, diminishing returns to scale to the R&D process imply a relation between these two cost functions as pictured in Figure 3.1. In terms of the model employed here this translates into

---

[3] For a recent survey on the issue of diminishing returns to scale in R&D see Patel and Pavitt [1995].

$$\frac{\partial C^{-1}[c_i(x_i)]}{\partial x_i} < 0. \tag{3.7}$$

and

$$\frac{\partial^2 C^{-1}[c_i(x_i)]}{\partial x_i^2} > 0. \tag{3.8}$$

Finally, it is assumed that all firms maximize their profits.

3.2.3 Government

The subsidy a firm receives depends only on its R&D investment and is given by $g(x_i,s)$, where s is government's control variable. It is assumed that

$$\partial g(x_i,s)/\partial x_i > 0. \tag{3.9}$$

This means that the R&D-subsidy is higher the more resources a firm devotes to innovative activity. Also it is assumed that

$$\partial g(x_i,s)/\partial s \geq 0 \quad \Leftrightarrow \quad x_i \geq 0, \tag{3.10}$$

which indicates that only innovating firms receive an R&D-subsidy. Moreover, $g(x_i,s)$ is assumed to be twice continuously differentiable in x_i.

Given the assumed structures of the market and the production process, it follows that firm i's second stage (before tax) profits are

$$\pi_i^r(p_i,q_i,x) = \{p_i(x)-c_i(x)\}q_i(x)-C(x_i)+g(x_i,s). \tag{3.11}$$

As in Chapter 2, firms are taxed in the product market to finance the R&D-subsidy. In equilibrium, after-tax profits are given by[4]

$$\pi_i^* = (1-t^*)\pi^{**} = (1-t^*)[\{p^*-c(x^*)\}q^*-C(x^*)+g(x^*,s^*)]. \tag{3.12}$$

where again stars denote equilibrium outcomes and where t^* is the equilibrium tax rate. The *NBP* in the framework employed here reads

$$NBP = tn\pi^*-ng(x^*,s^*), \tag{3.13}$$

which defines the tax rate as

---

[4] Observe that in equilibrium $s=s^*$, hence $x_i=x_j=x^*$, $p_i(x^*)=p_j(x^*)=p^*$, and $q_i(x^*)=q_j(x^*)=q^*$, $\forall i \neq j$.

$$t^* = \frac{ng(x^*,s^*)+NBP}{n\pi^{s^*}}.$$ (3.14)

In equilibrium, after tax profits thus equal

$$\pi_i^* = \left(1-\frac{ng(x^*,s^*)+NBP}{n\pi^{s^*}}\right)\pi^{s^*} = \{p^*-c(x^*)\}q^*-C(x^*)-\frac{NBP}{n}.$$ (3.15)

Also in this chapter we confine ourselves to a balanced budget policy. Hence, $NBP=0$, and the same observation as made in Section 2.2.3 applies.

The government's objective is to maximize net total surplus using it's control variable s, supposing that firms' output and R&D investments are the equilibrium outcomes of the appropriate market game. For the market described here, equilibrium consumer's surplus equals

$$CS^* = \sum_{i=1}^{n} CS_i^* = \sum_{i=1}^{n}\{V_i(q^*)-p^*q^*\} = V(q^*)-np^*q^*.$$ (3.16)

where $V_i(q)$ is gross surplus of commodity i as part of the consumption bundle q. Producer's surplus in equilibrium is given by

$$PS = n(1-t^*)\pi^{s^*}.$$ (3.17)

The sum over (3.16) and (3.17), net total surplus, is then given by

$$W = V[q^*]-n\{c(x^*)q^*(x^*)-C(x^*)\}.$$ (3.18)

Having outlined the structure of the market considered, the equilibrium of the model is to be determined. Observe that each firm's market behaviour is described by a two-stage game; first they simultaneously determine their R&D investment; given these, they simultaneously set either price or output, depending on the type of competition in the product market. It is only in the first stage of this game that firms are allowed to cooperate. In what follows the market equilibria under cooperative and noncooperative R&D are characterized.

3.3 Market equilibrium I: noncooperative R&D

3.3.1 Second-stage Cournot competition and noncooperative R&D

Maximizing (3.11) with respect to $q_i(x)$ gives the first-order condition[5]

$$p_i[q(x)] + \frac{\partial p_i[q(x)]}{\partial q_i(x)} q_i(x) - c_i(x) = 0,\tag{3.19}$$

which defines the equilibrium quantity, conditional on R&D investments, under second-stage Cournot competition

$$\hat{q}_{IC}(x),\tag{3.20}$$

where a hat refers to the conditional nature of the equilibrium, and where *I* refers to the noncooperative nature of the first stage of the game. Observe that (3.20) implicitly defines the second stage equilibrium price conditional on the R&D investments through (3.1).

First stage profits are then given by

$$\hat{\pi}_{iC}(x) = \{ p_i[\hat{q}_C(x)] - c_i(x) \} \hat{q}_{iC}(x) - C(x_i) + g(x_i,s).\tag{3.21}$$

Optimal levels of R&D investment are determined by maximising (3.21) with respect to x_i, which results in the following first-order condition (where use is made of (3.19))

$$\hat{q}_{iC}(x) \sum_{\substack{j=1 \\ j \neq i}}^{n} \left\{ \frac{\partial p_i[\hat{q}_C(x)]}{\partial \hat{q}_{jC}(x)} \frac{\partial \hat{q}_{jC}(x)}{\partial x_i} \right\} - \frac{\partial c_i(x)}{\partial x_i} \hat{q}_{iC}(x) - \frac{\partial C(x_i)}{\partial x_i} + \frac{\partial g(x_i,s)}{\partial x_i} = 0.\tag{3.22}$$

This defines the equilibrium under Cournot competition in case of noncooperative R&D, conditional on the R&D-subsidy

$$\{ \hat{x}_{IC}(s), \hat{q}_{IC}(s) \}.\tag{3.23}$$

The authorities are confronted with a welfare function equal to

$$\hat{W}_{IC}(s) = V(\hat{q}_{IC}[\hat{x}_{IC}(s)]) - nc[\hat{x}_{IC}(s)]\hat{q}_{IC}[\hat{x}_{IC}(s)] - nC[\hat{x}_{IC}(s)].\tag{3.24}$$

The optimal R&D-subsidy is then obtained by maximising (3.24) over *s*, associated with which is the following first-order condition

---

[5] Throughout Section 3.3 and Section 3.4 it is assumed that second order and stability conditions are satisfied.

$$\frac{\partial V(\hat{q}_{IC}[\hat{x}_{IC}(s)])}{\partial \hat{q}_{IC}[\hat{x}_{IC}(s)]} \frac{\partial \hat{q}_{IC}[\hat{x}_{IC}(s)]}{\partial \hat{x}_{IC}(s)} \frac{\partial \hat{x}_{IC}(s)}{\partial s}$$

$$- n \left\{ \frac{\partial c[\hat{x}_{IC}(s)]}{\partial \hat{x}_{IC}(s)} \frac{\partial \hat{x}_{IC}(s)}{\partial s} \hat{q}_{IC}[\hat{x}_{IC}(s)] \right. \tag{3.25}$$

$$\left. + \frac{\partial \hat{q}_{IC}[\hat{x}_{IC}(s)]}{\partial \hat{x}_{IC}(s)} \frac{\partial \hat{x}_{IC}(s)}{\partial s} c[\hat{x}_{IC}(s)] - \frac{\partial C[\hat{x}_{IC}(s)]}{\partial \hat{x}_{IC}(s)} \frac{\partial \hat{x}_{IC}(s)}{\partial s} \right\} = 0.$$

This defines the optimal R&D-subsidy under noncooperative R&D with Cournot competition in the product market,

$$s_{IC}^{*}. \tag{3.26}$$

3.3.2 Second-stage Bertrand competition and noncooperative R&D

If competition in the product market is over price, then in the second stage of the game (3.11) is maximised over $p_i(x)$. The corresponding first-order condition reads

$$q_i[p(x)] + \frac{\partial q_i[p(x)]}{\partial p_i(x)} \{p_i(x) - c_i(x)\} = 0, \tag{3.27}$$

and defines the equilibrium of the second stage game, conditional on R&D efforts, given that competition in the product market is over price,

$$\hat{p}_B(x). \tag{3.28}$$

Observe that (3.28) implicitly defines the conditional second stage equilibrium quantity through (3.2).
First stage profits then equal

$$\hat{\pi}_{iB}(x) = q_i[\hat{p}_B(x)] \{\hat{p}_{iB}(x) - c_i(x)\} - C(x_i) + g(x_i, s). \tag{3.29}$$

Optimal R&D investments are determined by maximising (3.29) over x_i, which yields the first-order condition (where (3.27) is used)

$$\{\hat{p}_{iB}(x) - c_i(x)\} \sum_{\substack{i=1 \\ i \neq j}}^{n} \left\{ \frac{\partial q_i[\hat{p}_B(x)]}{\partial \hat{p}_{jB}(x)} \frac{\partial \hat{p}_{jB}(x)}{\partial x_i} \right\} - \frac{\partial c_i(x)}{\partial x_i} q_i[\hat{p}_B(x)] - \frac{\partial C(x_i)}{\partial x_i} + \frac{\partial g(x_i, s)}{\partial x_i} = 0. \tag{3.30}$$

Implicitly this condition describes the equilibrium of the noncooperative game under second-stage

Bertrand competition conditional on the R&D-subsidy

$$\{\hat{x}_{IB}(s),\hat{p}_{IB}(s)\}. \tag{3.31}$$

Social welfare is again given by (3.24) with the appropriate price and quantities, as derived under second-stage Bertrand competition, inserted therein. Hence, (3.25) also determines the optimal R&D-subsidy under Bertrand competition, given that the R&D efforts are determined by (3.30),

$$\overset{\bullet}{s}_{IB}. \tag{3.32}$$

3.4 Market equilibrium II: cooperative R&D

As already observed in Chapter 2, allowing for cooperation in R&D does not affect the strategic decision of the production stage. That is, under Cournot competition the equilibrium of the second stage game is still (implicitly) defined by (3.20), while (3.28) is the (implicit) second stage game equilibrium when competition is over price.

3.4.1 Second-stage Cournot competition and cooperative R&D

In the first stage of the game, optimal R&D investments are determined by maximising *joint* profits. Under Cournot competition these are

$$\Pi_C[\hat{q}_C(x)] = \sum_{i=1}^{n} \pi_{iC}[\hat{q}_C(x)] = \sum_{i=1}^{n} \{\{p_i[\hat{q}_C(x)]-c_i(x)\}\hat{q}_{iC}(x)-C(x_i)+g(x_i,s)\}. \tag{3.33}$$

Maximising this expression with respect to x_k gives as first-order condition (where (3.19) is used)

$$\sum_{i=1}^{n}\left\{\left[\left[\sum_{\substack{j=1\\j\neq i}}^{n}\left\{\frac{\partial p_i[\hat{q}_C(x)]}{\partial \hat{q}_{jC}(x)}\frac{\partial \hat{q}_{jC}(x)}{\partial x_k}\right\}-\frac{\partial c_i(x)}{\partial x_k}\right]\hat{q}_{iC}(x)\right\}-\frac{\partial C(x_k)}{\partial x_k}+\frac{\partial g(x_k,s)}{\partial x_k}=0, \tag{3.34}$$

which defines the market equilibrium, conditional on the R&D-subsidy, when competition in the product market is over quantities and when firms cooperate in setting their R&D investment

$$\{\hat{x}_{IIC}(s),\hat{q}_{IIC}(s)\}. \tag{3.35}$$

Here *II* refers to the cooperative nature of the first stage of the game.

Observe that the technological spillover rate is assumed to be unaffected by the cooperative agreement. That is, firms form an industry-wide R&D-cartel. This assumption, also implicitly made by Suzumura [1992], will be maintained throughout this chapter.

Given this market outcome, government seeks to maximise

$$W_{II}^C(s) = V(\hat{q}_{II}^C[\hat{x}_{II}^C(s)]) - nc[\hat{x}_{II}^C(s)]\hat{q}_{II}^C[\hat{x}_{II}^C(s)] - nC[\hat{x}_{II}^C(s)], \qquad (3.36)$$

over its control variable, the R&D-subsidy. It does so by setting s according to the following first-order condition

$$\frac{\partial V(\hat{q}_{IIC}[\hat{x}_{IIC}(s)])}{\partial \hat{q}_{IIC}[\hat{x}_{IIC}(s)]} \frac{\partial \hat{q}_{IIC}[\hat{x}_{IIC}(s)]}{\partial \hat{x}_{IIC}(s)} \frac{\partial \hat{x}_{IIC}(s)}{\partial s}$$

$$- n \left\{ \frac{\partial c[\hat{x}_{IIC}(s)]}{\partial \hat{x}_{IIC}(s)} \frac{\partial \hat{x}_{IIC}(s)}{\partial s} \hat{q}_{IIC}[\hat{x}_{IIC}(s)] \right. \qquad (3.37)$$

$$+ \frac{\partial \hat{q}_{IIC}[\hat{x}_{IIC}(s)]}{\partial \hat{x}_{IIC}(s)} \frac{\partial \hat{x}_{IIC}(s)}{\partial s} c[\hat{x}_{IIC}(s)] - \left. \frac{\partial C[\hat{x}_{IIC}(s)]}{\partial \hat{x}_{IIC}(s)} \frac{\partial \hat{x}_{IIC}(s)}{\partial s} \right\} = 0.$$

The optimal R&D-subsidy under cooperative R&D with Cournot competition in the product market, which follows from (3.37), is given by

$$s_{IIC}^{\bullet}. \qquad (3.38)$$

3.4.2 Second-stage Bertrand competition and cooperative R&D

Joint first-stage profits under second-stage Bertrand competition equal

$$\Pi_B[\hat{p}_B(x)] = \sum_{i=1}^{n} \pi_{iB}[\hat{p}_B(x)] = \sum_{i=1}^{n} \{ \{ \hat{p}_{iB}(x) - c_i(x) \} q_i[\hat{p}_B(x)] - C(x_i) + g(x_i, s) \}. \qquad (3.39)$$

R&D investments are determined by maximising (3.39) with respect to x_k. This results in the following first-order condition (where (3.27) is used)

$$\sum_{i=1}^{n} \left\{ \left[\sum_{\substack{j=1 \\ j\neq i}}^{n} \left\{ \frac{\partial q_i[\hat{p}_B(x)]}{\partial \hat{p}_{jB}(x)} \frac{\partial \hat{p}_{jB}(x)}{\partial x_k} \right\} \right] [\hat{p}_{iB}(x) - c_i(x)] - q_i[\hat{p}_B(x)] \frac{\partial c_i(x)}{\partial x_k} \right\}$$

$$- \frac{\partial C(x_k)}{\partial x_k} + \frac{\partial g(x_k, s)}{\partial x_k} = 0. \qquad (3.40)$$

Conditional on the R&D-subsidy, the cooperative equilibrium under Bertrand competition of the product market is thus defined, and denoted by

$$\{\hat{x}_{IIB}(s), \hat{p}_{IIB}(s)\}. \tag{3.41}$$

The welfare function under Bertrand competition is also given by (3.36) with the appropriate quantities and R&D investments (that is, those resulting from second-stage Bertrand competition). Hence, (3.37) also determines the optimal R&D-subsidy when firms compete over price in the product market, given that the conditional equilibrium R&D efforts are determined by (3.40), and is given by

$$\dot{s}_{IIB}. \tag{3.42}$$

3.5 An Equivalence Theorem

Having characterized the market equilibria for both second-stage Cournot and Bertrand competition, as they emerge under competitive and cooperative R&D, the effects of the two R&D stimulating policies can be considered. To assess the consequences of government interventions in the market process, Suzumura [1992] proposes to use a second-best level of social welfare as benchmark. In particular, he considers the R&D investment which maximizes social welfare *given* the second stage market outcome. Suzumura [1992] justifies this measure, as opposed to the first-best solution in which the government both sets output or price and R&D investment (see e.g. d'Aspremont and Jacquemin [1988]), by observing that (Suzumura [1992, p.1308])

> the enforcement of the first-best arrangement may require considerable leverage on the government
> vis-à-vis private firms, something which may be hard to secure in reality. What is needed is an
> evaluation of the social gains from cooperative R&D within the alternative feasible arrangements.

Within the framework employed here this means that

$$W_C^{SP}(x) = V[\hat{q}_C(x)] - nc(x)\hat{q}_C(x) - nC(x), \tag{3.43a}$$

or

$$W_B^{SP}(x) = V[\hat{q}_B(x)] - nc(x)\hat{q}_B(x) - nC(x), \tag{3.43b}$$

should be maximised over x, depending on the type of product market competition. The concomitant first-order condition reads

$$\frac{\partial V[\hat{q}_T(x)]}{\partial \hat{q}_T(x)} \frac{\partial \hat{q}_T(x)}{\partial x} - n \left\{ \frac{\partial c(x)}{\partial x} \hat{q}_T(x) + \frac{\partial \hat{q}_T(x)}{\partial x} c(x) - \frac{\partial C(x)}{\partial x} \right\} = 0, \tag{3.44}$$

for $T = C,B$, where T stands for *type*. Label the optimal R&D levels under Cournot and Bertrand competition in the product market, implicitly defined by (3.44), x_C^{SP} and x_B^{SP} respectively. The following equivalence theorem can then be stated, the proof of which is given in appendix 3A.

Theorem 3.1

$$x_T^{SP} = \hat{x}_{IT}(s_{IT}^{\bullet}) = \hat{x}_{IIT}(s_{IIT}^{\bullet}), \text{ for } T = C,B.$$

The equivalence theorem states that optimally subsidizing an R&D-cartel (without cooperation in the product market) and optimally subsidizing noncooperative R&D, for which firms are taxed in the production market, results in exactly the same R&D investments. As a result, the complete market outcome and social welfare will be the same under the two subsidized regimes. In case of full spillovers the equivalence theorem indicates that subsidizing an RJV or an RJV-cartel leads to the same market outcome.

Observe that Theorem 3.1 is a generalization of Proposition 2.3. Indeed, a result obtained within the linear-quadratic framework carries over to a more general setting.

3.6 Cooperation versus subsidization under second-stage Cournot competition

Suzumura [1992] uses a more specific model to assess cooperative and noncooperative R&D than the one employed here. In particular, he assumes products to be homogeneous. Moreover, the costs of R&D are given by[6]

$$C(x_i) = x_i.$$

Finally, Suzumura [1992] considers only second-stage Cournot competition. For this setting he proves

---

[6] Suzumura [1992] says (Suzumura [1992, p.1308])

> Diminishing returns on R&D can be incorporated without affecting any of the results in this paper.

However, the fact that $C(x_i) = x_i$ does not imply that there are constant returns to R&D investments (as implied by the quote above), since what matters is the relation between the cost of investing in R&D and the extent to which R&D lowers marginal costs of production (recall Figure 3.1, (3.7) and (3.8)). Given that Suzumura does not specify $c_i(x)$, his analysis could be seen as one in which there are diminishing returns on R&D.

Theorem 3.2 (Suzumura)

$$x_C^{SP} > \hat{x}_{IIC}(0).$$

Theorem 3.2 indicates that the resources devoted to R&D in an R&D-cartel falls short of the second-best level, as defined by (3.44). As proved by Suzumura [1992], the theorem holds for all levels of spillovers. Hence, an RJV-cartel also devotes resources to R&D which are below those considered second-best.

Combining then Theorem 3.2 with the equivalence theorem leads to

Theorem 3.3

$$\hat{x}_{IC}(s_{IC}^*) > \hat{x}_{IIC}(0).$$

Assuming that the spillover rate is not influenced by the cooperative nature of the first stage of the game, Theorem 3.3 states that the provision of direct R&D-subsidies, for which firms are taxed in the product market, is more effective in promoting private R&D efforts than allowing firms to form an R&D-cartel. In case of full spillovers Theorem 3.3 indicates that an optimally subsidized RJV devotes more resources to R&D than a non-subsidized RJV-cartel.

3.7 Conclusions

The purpose of this short chapter is to generalize some of the key results presented in Chapter 2. Using a general framework it is shown that some important conclusions of the previous chapter are not artifacts of the linear-quadratic framework. In particular, the fact that optimally subsidized noncooperative R&D and optimally subsidized cooperative R&D lead to the same market outcome is shown to carry over to a general framework. Also, it is shown that for a more particular setting (although still more general than the one employed in Chapter 2), under second-stage Cournot competition, providing optimal R&D-subsidies is more effective in raising private R&D efforts than allowing firms to form R&D-cartels.

The policy conclusions drawn in Chapter 2 are reinforced by the analysis of the present chapter. In particular, it remains doubtful that allowing firms to cooperate in their research is the appropriate way to encourage private R&D efforts, knowing that a more effective instrument is at hand which does not trigger the threat that firms will extend the collusive agreement to the product market, a type of collusive behaviour that is socially undesirable but difficult to monitor.

However, substantial research is still to be conducted to assess the two R&D stimulating policies more thoroughly. R&D-subsidies especially should be analyzed in more detail. For instance, the objections against R&D-subsidies, as spelled out in Section 1.6, should receive more consideration in formal models. Also, much more empirical research is needed to determine adequately the precise effects of R&D-subsidies. Indeed, the analyses presented in Chapters 2 and 3 are just a first step towards understanding the economic implications of a widely used policy portfolio for stimulating private R&D: subsidizing R&D-cooperatives.

Appendix 3A Proof of Theorem 3.1

Observe that (see (3.25), (3.37) and (3.44))

$$\frac{\partial W_{IT}[\hat{x}_{IT}(s)]}{\partial s} = \frac{\partial W_{IT}[\hat{x}_{IT}(s)]}{\partial \hat{x}_{IT}(s)} \frac{\partial \hat{x}_{IT}(s)}{\partial s} = \frac{\partial W_T^{SP}(x)}{\partial x} \frac{\partial \hat{x}_{IT}(s)}{\partial s},$$ (3A.1)

and

$$\frac{\partial W_{IIT}[\hat{x}_{IIT}(s)]}{\partial s} = \frac{\partial W_{IIT}[\hat{x}_{IIT}(s)]}{\partial \hat{x}_{IIT}(s)} \frac{\partial \hat{x}_{IIT}(s)}{\partial s} = \frac{\partial W_T^{SP}(x)}{\partial x} \frac{\partial \hat{x}_{IIT}(s)}{\partial s},$$ (3A.2)

for $T = C, B$. Therefore

$$\frac{\partial W_{IT}[\hat{x}_{IT}(s)]}{\partial s} = 0 \quad \Leftrightarrow \quad \frac{\partial W_T^{SP}(x)}{\partial x} = 0,$$ (3A.3)

and

$$\frac{\partial W_{IIT}[\hat{x}_{IIT}(s)]}{\partial s} = 0 \quad \Leftrightarrow \quad \frac{\partial W_T^{SP}(x)}{\partial x} = 0,$$ (3A.4)

for $T = C, B$. Hence in equilibrium it must be that

$$\hat{x}_{IT}(s_{IT}^{\bullet}) = x_T^{SP} = \hat{x}_{IIT}(s_{IIT}^{\bullet}),$$ (3A.5)

for $T = C, B$.

Q.E.D.

Part II

Product Differentiation

In fact, as regards the more important knowledge, I do believe
she is invariable superficial. The depth lies in the valleys where
we seek her, and not upon the mountain-tops where she is found.

- Auguste Dupin -

4   Spatial competition with an outside good

4.1 Introduction

Space has intrigued economists as long as there has been an economics profession. Economists like Johann von Thünen and Léon Walras considered space in some of their writings and nowadays contributions to economic theory that explicitly consider the role of space abound.

Many analyses consider the question where competing firms in a spatial market will locate their production facility. On a micro-economic level a fundamental breakthrough in this area of research was established in 1929 by Harold Hotelling.[1] In an attempt to obtain stability in competition in a Bertrand-like world, he developed a stylized model in which two firms compete not only in price but also in space. In particular, both merchants first simultaneously set up shop, after which they compete in price. Locations can be chosen along a line "which may be Main Street in a town or a transcontinental railroad" (p.45). Solving this two-stage game backwards (!) Hotelling came to the belief that both merchants will locate back-to-back at the centre of the market: the famous Principle of Minimum Differentiation.

Although there are some problems with Hotelling's analysis[2], he did succeed in putting space on micro-economists' research agendas. Indeed, much of the research being done in this area is based on Hotelling's seminal contribution. Moreover, as Hotelling realized, his visualization of space has a perfect analogy in the extent to which products are differentiated. That is, firms' location can be thought of as their position in the variety plane if Main Road is interpreted as the dimension along which consumers' preferences are distributed.

Phlips and Thisse [1982] point out that Hotelling's model only applies to what we now call *horizontally* differentiated products (Phlips and Thisse [1982, p.2]):

> Differentiation is said to be *horizontal* when (..) between two products the level of some characteristics is augmented while it is lowered for some others, as in the case of different versions (..) of a car. (..) [A consumer] will buy the "closest" product in terms of a certain *distance* function. (..) Differentiation is called *vertical* when (..) between two products the level of all characteristics is augmented or lowered, as in the case of cars of different series. (..) There is unanimity to rank the products according to a certain *order*.

According to Hotelling's principle then, products will be horizontally differentiated to a minimum extent, leaving consumers with a narrow choice of varieties.

As will become clear below, Hotelling's analysis contains a fundamental flaw which makes it impossible to say anything as to where both merchants locate. Since the identification of this flaw

---

[1] Whenever in this chapter we refer to Hotelling, it is to his famous "Stability in Competition" article of 1929.

[2] Hotelling's model and the problems associated with it will be discussed in detail in the next section.

in 1979 by d'Aspremont, Gabszewicz and Thisse, many studies have relaxed some of the (implicit) assumptions made by Hotelling. Several of these studies will be reviewed below. Meanwhile we can state the rather indecisive conclusion that comes from this literature: the price-location equilibrium depends very much on which assumption is relaxed. Because to date no analysis has been presented in which all of Hotelling's assumptions are relaxed, no definite solution as to where firms will locate their plant is as yet provided by economic theory.

The bulk of this chapter is concerned with yet another departure from Hotelling's model, a departure which has not been addressed adequately before. It contains the description of a spatial duopoly in which consumers have an upper-limit on the price they are willing to pay for the horizontally differentiated product. For this generalization of Hotelling's model his principle does not hold. Rather, the analysis leads us to conclude that if firms compete, in most cases they will locate at the market quartiles, implying a Principle of Almost Intermediate Differentiation.

4.2 The Principle of Minimum Differentiation conceived and flawed

As mentioned above, Hotelling visualized the market as a line of length l along which consumers are uniformly distributed. In this market two merchants, a and b, simultaneously set up shop (stage one), after which they simultaneously determine their mill price (stage two). Fixed and marginal cost of production are assumed to be zero. Because producers adhere to free-on-board mill pricing, consumers bear the cost of transportation. This is assumed to be the same for each unit of distance; that is, the transport rate, t, is constant. Each consumer always purchases one unit of the (homogeneous) good of either firm, which means that demand is completely inelastic. All consumers enjoy the same gross surplus of utility when buying this good, denoted by u. Net utility of a consumer located at h^* thus equals

$$U(h^*, h_i) = u - t |h^* - h_i| - p_i, \qquad i = a, b, \tag{4.1}$$

where p_i is the mill price set by merchant i. The marginal consumer, M, is indifferent between buying from firm a or b. His location thus marks the dividing line between the two firms' market segments and is implicitly defined by

$$p_a + t |M - h_a| = p_b + t |l - M - h_b|. \tag{4.2}$$

Figure 4.1 illustrates this type of market configuration (ignore for the moment the bold dashed lines). Observe that firm a's location, h_a, is defined as the distance from the left end of the market while that of firm b, h_b, is defined as the distance from the right end of the market.
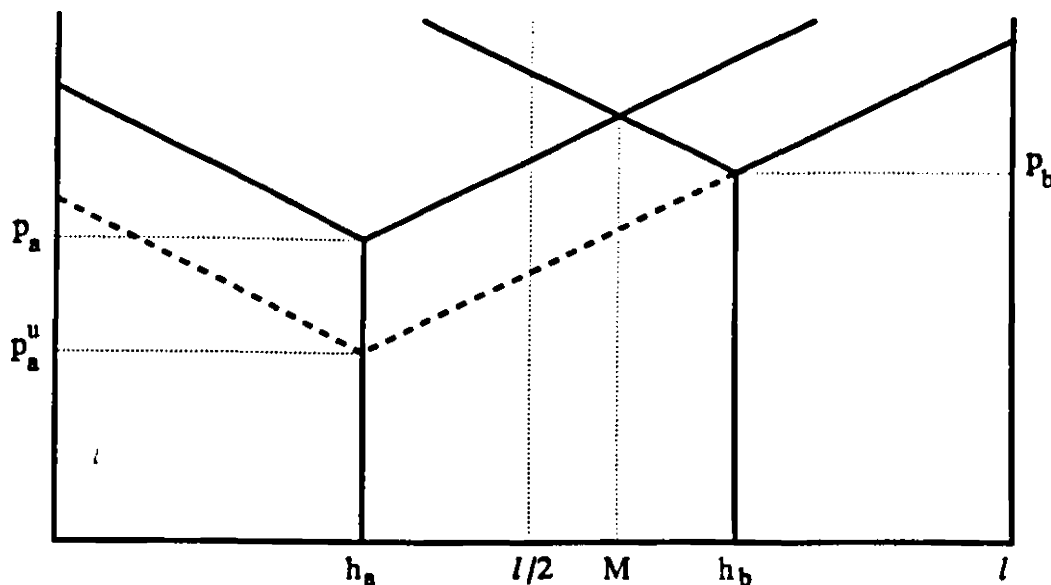
Figure 4.1 Hotelling's spatial model

For this type of market, second stage profits of firm i equal

$$\pi_i(p_i,p_j;h_i,h_j) = \begin{cases} p_i l, & p_i < p_j - t(l - h_i - h_j), \\ p_i\left(\dfrac{p_j - p_i}{2t} + \dfrac{l + h_i - h_j}{2}\right), & |p_i - p_j| \le t(l - h_i - h_j), \\ 0, & p_i > p_j + t(l - h_i - h_j), \end{cases}$$

(4.3)

$i,j,=a,b,\ i \ne j$ (to avoid the cumbersome notation $i,j=a,b,\ i \ne j$ this is implicit throughout this chapter). The first part of equation (4.3) corresponds to the situation in which firm i undercuts firm j (see the bold dashed lines in Figure 4.1): the consumer located at firm j's location perceives a lower effective price from firm i, $p_i + t(l - h_i - h_j)$, than from firm j, p_j. Because transportation cost is linear the same is true for all consumers located in firm j's hinterland. In this situation firm i serves the whole market while firm j is confronted with zero demand. The third part of firm i's profit function refers to the case in which firm j undercuts firm i, leaving the latter with no demand.

Hotelling only considered the case in which both merchants serve part of the market, that is the middle part of equation (4.3). This led him to derive an alleged equilibrium price which is in fact *not* an equilibrium price as he defined 'equilibrium' for his spatial model. But before we examine this problem in more detail, let us follow Hotelling's reasoning first.

Maximizing the middle part of (4.3) with respect to p_i and solving for the symmetric solution induces firm i to quote a price equal to

$$p_i^*(h_i, h_j) = t\left(l + \frac{h_i - h_j}{3}\right).$$

(4.4)

Concomitant first stage profits then equal

$$\pi_i^*(h_i, h_j) = \frac{t}{2}\left(l + \frac{h_i - h_j}{3}\right)^2.$$

(4.5)

It is then immediate that in the first stage both firms want to enlarge their respective hinterlands as much as possible, implying that in a symmetric equilibrium both merchants are located back to back at the market centre. This is the famous Principle of Minimum Differentiation.

Hotelling claimed to have identified a general phenomenon when he writes (p.54)

> The mathematical analysis thus leads to an observation of wide generality. Buyers are confronted everywhere with an excessive sameness. When a new merchant or manufacturer sets up shop he must not produce something exactly like what is already on the market or he will risk a price war of the type discussed by Bertrand in connection with Cournot's mineral spring. But there is an incentive to make the new product very much like the old, applying some slight change which will seem an improvement to as many buyers as possible without ever going far in that direction. The tremendous standardisation of our furniture, our houses, our clothing, our automobiles and our education are due in part to the economies of large-scale production, in part to fashion and imitation. But over and above these forces is the effect we have been discussing, the tendency to make only slight deviations in order to have for the new commodity as many buyers of the old as possible, to get, so to speak, *between* ones' competitors and a mass of customers.

He continues (p.54)

> So general is this tendency that it appears in the most diverse fields of competitive activity, even quite apart from what is called economic life.

and then discusses how similar the American Republican and Democratic parties are. Finally, he concludes (p.57)

> Our cities become uneconomically large and the business districts within them are too concentrated. Methodist and Presbyterian churches are too much alike; cider is too homogeneous.

4.2.1 Instability in competition

Yet, economic intuition contradicts Hotelling's conclusion. If firms are located at the same place we have essentially the situation pictured by Bertrand, although in Hotelling's version both firms enjoy positive profits! Fifty years after Hotelling published his influential article a fundamental flaw in its analysis was revealed. D'Aspremont et al. [1979] rigorously examine the above mentioned possibility of undercutting and derive conditions under which Hotelling's equilibrium price is valid. In particular they show that firm i's maximization problem of the second stage leads to

$$p_i^*(h_i, h_j) = t\left(l + \frac{h_i - h_j}{3}\right) \quad \Leftrightarrow \quad \left(l + \frac{h_i - h_j}{3}\right)^2 \geq \frac{4l(h_i + h_j)}{3}, \tag{4.6}$$

if $h_a + h_b < l$, while $p_i^*(h_i, h_j) = 0$, if $h_a + h_b = l$. The condition in (4.6) under which Hotelling's equilibrium price is valid follows from equating competitive profits with those obtained when supplying the whole market. For example, if firm a wants to capture the whole market it has to quote a price equal to $p_a^u(h_a, h_b) = p_b^*(h_a, h_b) - t(l - h_a - h_b)$ (see Figure 4.1). Its profits then equal $p_a^u l$ and exceed $\pi_a^*(h_a, h_b) = t[l + (h_a - h_b)/3]^2/2$ whenever the condition in (4.6) is not met.

Observe that for symmetric locations undercutting becomes profitable if both firms are located inside the market quartiles and nothing can be said about the tendency to agglomerate. *If* on the other hand both merchants locate at the same place (for symmetric locations that being the market centre) the situation described by Bertrand arises and neither firm is able to keep its price above marginal cost, much less earn positive profits.

Hotelling does seem to have been aware of the possibility of a Bertrand-like price war, as can be inferred from the long quote above. Moreover, he explicitly considers the undercutting price when he writes (p.45)

> Now b's price may be higher than a's, but if b is to sell anything at all he must not let his price exceed a's by more than the cost of transportation from a's place of business to his own. In fact he will keep his price p_b somewhat below the figure $p_a - t(l - h_a - h_b)$ at which a's goods can be brought to him.

He failed to realize however that when firms are located too closely together this strategy will drive prices and hence profits down to zero.

More generally, Hotelling did not distinguish the two effects which determine firms' behaviour: the strategic or market effect and the competitive or price effect. The strategic effect drives firms toward each other because both want to gain market share. But the competitive effect induces firms not to locate too closely together in order to avoid stronger price competition. In Hotelling's model the strategic effect dominates first, inducing both firms to move towards each other. But then the competitive effect destroys the equilibrium because Bertrand-like undercutting becomes profitable. It is the strategic effect Hotelling built his analysis on, ignoring the destructive force of price competition.

4.2.2 Non quasi-concavity

Four years after d'Aspremont et al. [1979] revealed the fundamental flaw in Hotelling's analysis, the same authors gave the decisive blow (see d'Aspremont et al. [1983]). They show that under Hotelling's assumptions the Principle of Minimum Differentiation *never* holds. If firms locate too closely together then either i) profits go to zero, or ii) a Nash-Cournot equilibrium in prices does not exist.

More generally, as observed by Gabszewicz and Thisse [1986], the non-existence of an equilibrium is due to the non-quasi concavity of profit functions. In Figure 4.2 firm a's first stage profits, as stated in (4.3), are drawn as a function of p_a for given locations while keeping firm b's price fixed at $\bar{p}_b$. Clearly this profit function is not quasi-concave. Anderson [1987] carefully constructs price-reaction functions for Hotelling's spatial duopoly and shows that these are discontinuous and do not intersect (see also Martin [1993, p.263-71]).

One obvious way to restore the existence of an equilibrium is to restrict either firm's action-set to the quasi-concave part of the profit functions. As will be shown in Subsection 4.6.1, the equilibrium that emerges depends crucially on *how* firms are restricted in their actions. We first proceed however with some other digressions from Hotelling which restore the price-location equilibrium.

4.3 The Principle of Minimum Differentiation restored

Although Hotelling explicitly introduced a new dimension into the theory of price competition, it did not lead to stability in competition. As Phlips and Thisse [1982] remark (p.2-3)

> Hotelling did not depart far enough from the other assumptions that underpin the perfect competition and the Cournot-Bertrand models, to come to grips with the phenomenon of price discrimination

As mentioned in the introduction, by now there are numerous studies restoring the existence of an equilibrium in a spatial duopoly by, indeed, 'departing enough' from the Cournot-Bertrand models (that is, relaxing one or more assumptions (implicitly) made by Hotelling). A couple of these studies restore the Principle of Minimum Differentiation. In this subsection we will review some of these, followed by a discussion in the next subsection of analyses which claim the opposite result: the Principle of Maximum Differentiation.[3]

---

[3] The literature on spatial competition is too massive to completely review here. A result of the abundance of this literature is that survey articles also consider only part of the literature (see e.g. Graitson [1982]). See for other surveys e.g. Beckmann and Thisse [1986], Gabszewicz et. al [1986], Beath and Katsoulacos [1991], Anderson et. al [1992] and Gabszewicz and Thisse [1992].
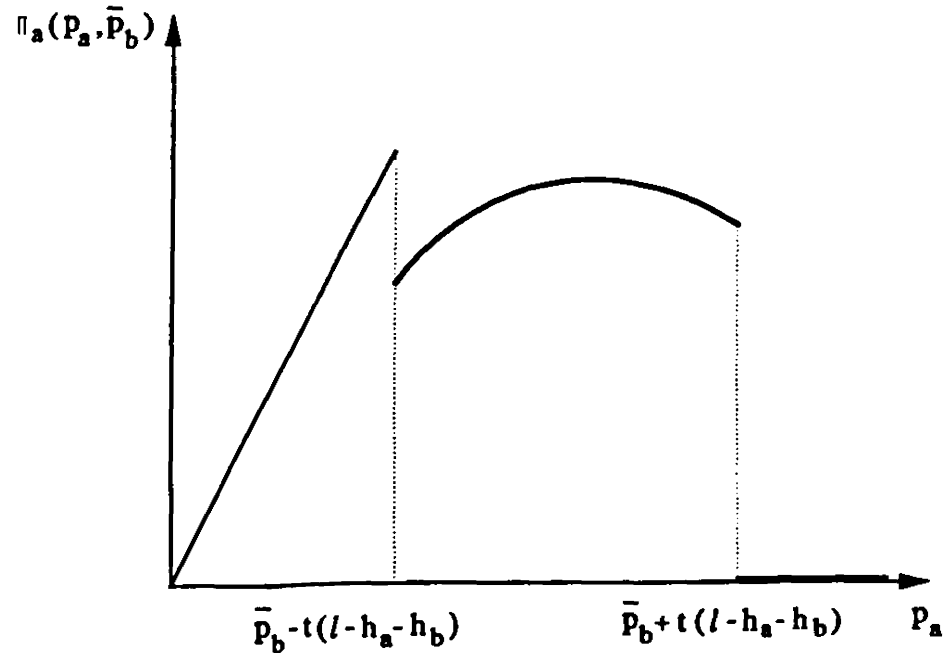
$$\Pi_a(P_a, \bar{P}_b)$$

[A figure showing firm a's profit curve rising linearly then, after a jump, forming a concave hump, plotted against P_a. The x-axis shows two marked values.]

$$\bar{P}_b - t(l - h_a - h_b) \qquad \bar{P}_b + t(l - h_a - h_b) \qquad P_a$$

Figure 4.2 Firm *a*'s second-stage profits in Hotelling's spatial duopoly

4.3.1 Cournot competition

Anderson and Neven [1990] show that when firms compete over quantities, after they have selected locations, and supposing that transport cost are convex, the Principle of Minimum Differentiation does hold. Indeed, existence and uniqueness of the equilibrium are thoroughly proved. In their model each point in space constitutes a particular linear demand. Hence, each firm will supply all points in space, although a firm's market share is largest at those points closest to its location. The tendency to agglomerate is then intuitively clear. Observe first that as long as a firm is not located at the market centre, it's hinterland is smaller than half the total market. Moving then towards the centre results in some loss in demand from a firm's hinterland, but this loss is smaller than the increase in demand due to the larger number of consumers the firm approaches. In addition, given that transport cost is convex, the rise in profits due to increased demand from the 'far' side of the market is more rapid than the loss in hinterland-profits. Assuming both firms follow this reasoning means that both locate at the centre of the market.

4.3.2 Repeated collusion

Friedman and Thisse [1993] consider a situation in which the two duopolists first determine non-cooperatively where to locate and then collude in setting their price. At the beginning of the game locations are chosen and are fixed forever while prices are set in each consecutive time period for an

infinite number of periods. To sustain repeated cooperation a trigger strategy is developed. Payoffs are such that the profit ratio of the partial collusive (repeated game) outcome is positively related to the same ratio of the single-shot equilibrium. In addition a reservation price is introduced which, however, is assumed to be sufficiently large that all consumers face a full price below this reservation price. The third and final deviation from Hotelling's original setup is that transportation cost is assumed to be quadratic. In a non-cooperative single-shot setting this leads firms to differentiate as much as possible (see d'Aspremont et al. [1979]; non-linear transport cost will be discussed in detail in Subsection 4.4.1). On the other hand, the single-shot full cooperation setting, with quadratic transportation cost, would lead firms to locate at the market quartiles. Friedman and Thisse [1993] show rigorously that in their context the unique and symmetric equilibrium has firms located at the market centre. Discussing this result Friedman and Thisse remark (p.642)

> Price competition was enough to destroy Hotelling's belief that spatial competition should result in 'excessive sameness'; price collusion is enough to support it.

4.3.3 Price-taking firms

If prices are regulated and firms consequently can only choose where to locate, the tendency to agglomerate characterizes the first stage of the game. Anderson and Engers [1994] provide details concerning existence and uniqueness of the equilibrium. The intuition of this result is clear. If price is fixed each duopolist will choose a location between its competitor and the major part of the market. If both firms follow this reasoning they will both locate at the median. Anderson and Engers [1994] show that demand must be 'sufficiently' inelastic for the agglomeration result to hold. The underlying logic for this restriction is similar to the argument put forward by Anderson and Neven [1990]. Moving away from the market centre will lead to some loss of consumers, namely those located between the moving firm and its competitor. However, the firm will be closer to all those customers it retains. Then, if demand is elastic enough, these consumers will put upward pressure on the moving firms' demand, which exceeds the loss in demand at the other side of the market. If on the other hand demand is inelastic enough the loss in demand due to moving away from the market centre dominates. Hence, in that case firms agglomerate.

A simple example shows that this result holds for any type of invertible transportation cost if demand is completely inelastic. Suppose that transportation cost are given by $f(\cdot): \mathbf{R}^+ \to \mathbf{R}^+$, with $f(0)=0$. The marginal consumer's location, M, is then implicitly given by

$$p_a + f(M - h_a) = p_b + f(l - M - h_b). \tag{4.7}$$

Since firms are price takers we have that $p_a = p_b = p$. Hence

$$f(M-h_a) = f(l-M-h_b), \tag{4.8}$$

or

$$M = (l+h_a-h_b)/2. \tag{4.9}$$

Firms a and b's first-stage profits are then given by[4]

$$\pi_a = p(l+h_a-h_b)/2, \tag{4.10}$$

and

$$\pi_b = p(l+h_b-h_a)/2, \tag{4.11}$$

respectively. From these it is readily derived that both firms want to enlarge their respective hinterlands as much as possible. Hence, they agglomerate at the market centre.

4.3.4 Sufficient heterogeneity

De Palma et al. [1985] assume that firms are differentiated by inherent characteristics. These differences are recognized by consumers, inducing them, in the perception of producers, to choose randomly between either merchant. Hence, firms cannot predict consumer's tastes. As a consequence, each seller endows each consumer with a probabilistic choice rule, the outcome of which cannot be predicted accurately. However, at the aggregate level producers are able to make predictions which correspond perfectly to observed aggregate behaviour. All this is modeled by reformulating (4.1) as

$$U(h^*,h_i) = u-t|h^*-h_i|-p_i+k\varepsilon^*, \tag{4.12}$$

where k is a constant capturing the heterogeneity between firms (the larger k, the more heterogeneous firms are), and where ε^* is a random variable with zero mean and unit variance. De Palma et al. [1985] assume each of these 'errors' to follow an independent Weibull distribution, implying that the probability that a consumer buys from a particular merchant is captured by a logit model. It is this probability firms can observe (that is, they can predict aggregate consumer behaviour perfectly).

Figure 4.3 displays the probability that a consumer will buy from firm a, p_r^a, for different values of k. In the extreme case of k being equal to 0 we are back in Hotelling's world and buying probabilities are given by the dashed kinked line: all consumers to the left of M buy from merchant a with probability 1 (see also Figure 4.1) while those located to the right of M always go to firm b. In the other extreme case, k approaching infinity, firms are so heterogeneous that a consumers' most preferred variety becomes irrelevant: each firm will be visited with equal probability (depicted in Figure 4.3 by the straight horizontal dashed line). For all values of k in between these two extremes

---

[4] Profits are always first-stage, since by assuming firms to be price-takers the second stage effectively disappears.
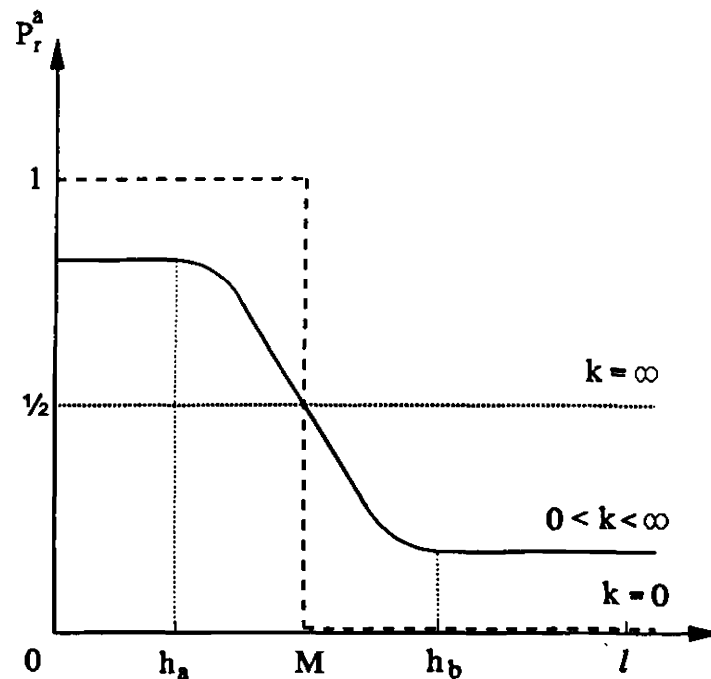
Figure 4.3 Probability of buying from merchant *a* when utility is random

consumers buy from the nearest firm with a probability p_r, which is between ½ and 1. The other firm will be visited with a probability of $1 - p_r$. For all consumers located in a firm's hinterland this probability is the same and constant since the other firm will always be further away. However, for consumers located in between the two firms this probability depends on the consumers' location relative to that of the two firms (see Figure 4.3).

As shown by de Palma et al. [1985], an important implication of introducing heterogeneity is that each firm's profit function is continuous. Hence, the problems associated with Hotelling's original analysis do not arise.

Analyzing then firms' consecutive pricing and location decisions, de Palma et al. [1985] show that the symmetric and unique equilibrium is firms clustering together at the market centre, given sufficient heterogeneity. In particular it must be that $k \geq tl$. De Palma et al. [1985] conclude (p.779):

> ...even though strong competition lessens the equilibrium market price under agglomeration, it may not render it low enough to overcome the advantage of higher market share at the centre. This happens whenever *k* is larger than *tl*, thus restoring Hotelling's Principle of Minimum Differentiation.

4.4 The Principle of Maximum Differentiation

Hotelling's initiation of the literature on 'excessive sameness' has its counterpart in studies resolving in (excessive) differences. It is to some of the explanations for this Principle of Maximum Differentiation that we now turn.

4.4.1 Non-linear transportation cost

When transportation cost is quadratic consumers have a stronger preference for a particular variety than in case of linear transportation cost because deviations from the most preferred variety are punished more in terms of utility loss. Hence, competition in the neighbourhood of a most preferred variety is expected to be stronger. That is, the competitive effect is stronger than in case of linear transportation cost. As shown by d'Aspremont et al. [1979], treating transportation cost in Hotelling's original model as being quadratic leads firms to disperse as much as possible. The fear of increased competition due to agglomeration is stronger than the accompanying gain in market share. Indeed, the competitive effect now dominates the strategic effect.[5]

Neven [1985] provides an elaborate description of the terse exposition of d'Aspremont et al. [1979]. He emphasizes that Hotelling's spatial model only has its analogy in *horizontally* differentiated products, referring to Phlips and Thisse [1982]. Interpreting Hotelling's analysis then as one describing horizontal product differentiation, Neven [1985] conjectures (p.320):

> The quadratic form, however, seems more natural than the linear one when consumers "move" in
> a product space. At least, it seems reasonable to assume that the marginal disutility of "moving"
> on the product line is increasing.

Hence, maximum differentiation occurs when firms want to horizontally differentiate their products.

But there is of course a continuum of utility loss functions between the linear and quadratic one embodying an increasing "marginal disutility of moving on the product line". Economides [1986] considers transportation cost of the type $|h^*-h_i|^\gamma$, where $1 \leq \gamma \leq 2$.[6] He shows that a unique (symmetric) price-location equilibrium exists for $1.26 \approx \bar{\gamma} < \gamma$. In particular, locations are at the corners of the market for $5/3 \leq \gamma \leq 2$, while they are given by $h^* = 5/4 - 3\gamma/4$ for $\bar{\gamma} < \gamma < 5/4$. For values of γ below the threshold $\bar{\gamma}$ no price-location equilibrium (in pure strategies) exists. In Figure 4.4 the locus of optimal locations is depicted by the thick solid line. The shaded area refers to the region in the variety space where no price-location equilibrium exist. Observe that for $\gamma < 1.26$ the optimal location lies within this area (as pictured by the dashed line).

---

[5] It would be very interesting to examine whether this result still holds when, to a certain extent, consumers' utility is random (as proposed by de Palma et. al [1985]).

[6] However, the length of the market and the transport rate are both normalized to be equal to 1 (Neven [1985] also treats the market as being of unit length). Relaxing these two assumptions might yield quit different outcomes than those reported by Economides [1986] as will become clear in the sequel of this chapter
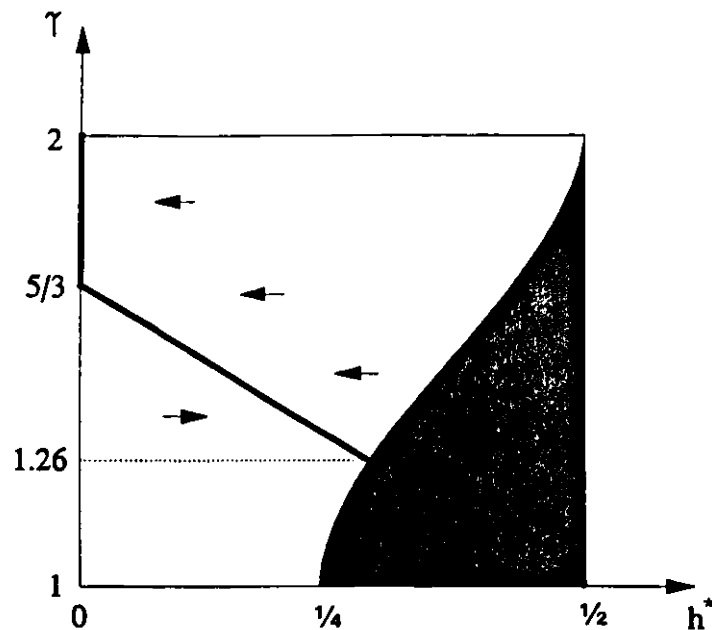
Figure 4.4 Locus of optimal locations if marginal disutility is increasing with distance

Economides' analysis shows that it is not enough to assume that marginal disutility is increasing with the distance between a consumers' most preferred variety and those offered for the Principle of Maximum Differentiation to hold. Indeed, for $1.26 < \gamma < 5/4$ transportation cost is increasing with distance but differentiation is not at its maximum.[7]

4.4.2 Reservation price

Another key assumption in Hotelling's analysis is that demand is perfectly inelastic. Lerner and Singer [1937] were the first to question this assumption when they write (p.148)

> ...it is only possible for each buyer to purchase one unit, irrespective of price, if there is no upper limit to his expenditure. In such a case the location of a single seller is indeterminate, for in any location a unit can be sold at an infinite price. It is necessary, therefore, to assume an upper limit to the price each buyer is willing to pay for his unit of the commodity if we are going to be at all realistic...the value of this limit is of primary importance for the problem of determining the location, and output of the producers.

---

[7] Gabszewicz and Thisse [1986] consider linear-quadratic transportation cost. They examine the existence of an equilibrium in a spatial duopoly when transportation cost takes the form $t_1 |h^* - h_i| + t_2(h^* - h_i)^2$. In a condensed appendix they show that no price equilibrium exists if the distance between the two firms, d, is too small. In particular, for a price equilibrium to be well defined it must be that $d \geq \min \{ 1/3, \sqrt{t_1}/(2\sqrt{t_2} + \sqrt{t_1}) \}$, given that the length of the market is normalized to equal 1 (see also Anderson [1988]).

Lerner and Singer [1937] challenge the analysis of Hotelling even more (p.151)

> He (Hotelling) assumes that, in choosing his location, *b* considers that *a*'s price will be that which is reached after a long process of price changes that follow from each producer adjusting himself each time the other changes his price, each producer choosing his own price each time on the assumption that the other's price is fixed. This means that, in choosing the location of his plant, *b* acts with greater wisdom than in fixing his price, and that once the location is fixed the greater wisdom evaporates. If *b* is no wiser in the one case than in the other (which is our assumption), he will take both *a*'s location and his price as fixed in choosing his own location and price.

They then re-examine Hotelling's model for several numbers of producers while assuming that each consumer is endowed with the same, fixed and finite *reservation* price (as opposed to Hotelling who implicitly assumed an infinitely high reservation price) and that both price and location are set simultaneously. The two-player game analyzed by Lerner and Singer [1937] however does not have a Nash-Cournot equilibrium, plagued as they are by the same difficulties Hotelling failed to address properly.

It took a good forty years before the line of research initiated by Lerner and Singer [1937] was taken up again.[8] In a seminal paper Salop [1979] analyzes a model of spatial competition in which an *outside* good is considered (which implicitly means that an upper limit is set to the price a consumer is willing to pay, as will become clear below). He pictured the market as a line of infinite length or a circle, where *n* firms have to establish their plant location and to set price. Salop [1979] abstracts from the decision as to what extent firms want to differentiate their products, since he *assumes* a symmetric distribution of plants over the spatial plane.

Analyzing this framework, Salop [1979] distinguishes three regions in a firm's demand function: (i) monopoly, (ii) competitive and (iii) super-competitive. These three possibilities are illustrated in Figure 4.5 for firm *a* (assuming firm *b*'s price to be fixed and equal to p_b) in case of only two goods. The reservation price is denoted by v, the monopoly price by p_a^m, the competitive price by p_a^c, and the supercompetitive price by p_a^u. Supercompetitive behaviour corresponds to undercutting, while competitive pricing resembles the case analyzed by Hotelling. Due to the existence of a reservation price each merchant has also the possibility to quote a price such that it acts as a local monopolist. The first remarkable result of Salop [1979] is that the elasticity of demand is higher in the monopoly price-region then in the competitive price-region. Further, observe that if firm *a* starts to lower its monopoly price a situation arises in which the reservation price exactly equals the full price the marginal consumer has to pay when buying from either producer. At this point the monopoly price-region touches the competitive price-region. Due to the differences in the elasticity of demand for these two regions a kink in the demand function occurs. Salop [1979] analyzes in detail the equilibrium at this kink and comes to his second remarkable result, namely that prices *fall* with an increase in either fixed or marginal cost. Salop [1979] remarks (p.149)

---

[8] Another way of eliminating completely inelastic demand is to endow each consumer with an inverse demand function, as proposed by Smithies [1941]. See Eaton [1972] for mathematical details.
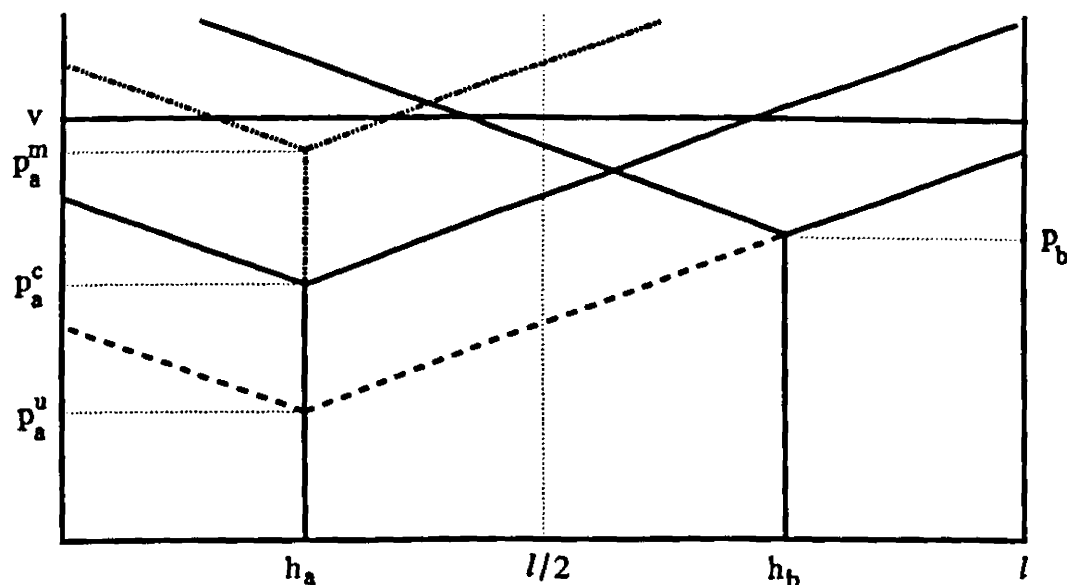
Figure 4.5 Monopoly, competitive and supercompetitive pricing-regions

This is a very striking result. If the increase in costs is interpreted as an excise tax levied on the industry, then the incidence of that excise tax is negative at the kinked equilibrium.

Welfare properties of the kinked equilibrium are also perverse. Indeed, an increase in cost which lowers prices can improve welfare.

As mentioned above, Salop [1979] explicitly assumes that firms are symmetrically located along the circle. That is, firm's locations are treated as exogenous. Economides [1989] generalizes Salop's analysis by introducing a genuine location-stage. Contrary to Salop [1979] however, Economides [1989] treats transportation cost as quadratic in distance. Under symmetry assumptions regarding the distribution of consumer's preferences the subgame-perfect price-location equilibrium entails firms locating equidistantly on the circle. Moreover, as in Salop [1979], Economides [1989] shows that in his setup the price subgame has three candidate equilibria (depending on the reservation price and the marginal cost functions): competitive, kinked, or monopolistic.

An approach closer to Hotelling's original setup is that of Economides [1984]. He analyzes the consequences for both price and location of introducing a finite reservation price in Hotelling's spatial duopoly. In that case, both merchants tend to move away from each other until they are local monopolists, thereby violating the Principle of Minimum Differentiation. Economides [1984] observes

The essential reason for this result is that, for relatively not-too-high reservation prices, firms are not guaranteed the purchases of the consumers who are located near the edges of the market.

However, this only holds if the market is large enough to host two monopolists. As will become clear below, Economides [1984] only considers a *low* reservation price relative to the length of the market and the transportation cost. That is, to date a full description of Hotelling's spatial duopoly with a finite reservation price has not been presented.

4.5 Minimum or maximum differentiation?

The previous two sections have shown that the literature on spatial duopolies does not provide (yet) a unanimous answer as to where two price-competing firms will locate. Due to the work of d'Aspremont et al. [1979] we do know that Hotelling was still too close to Bertrand in order to say something about locational choices. But, depending on the direction in which one departs from Hotelling's original setup, either it is predicted that firms differentiate to a minimum or to some extent, or possibly to a maximum.

The sequel of this chapter is devoted to yet another departure from Hotelling, albeit a modest one. It contains the entire analysis of Hotelling's model when a finite reservation price is introduced and can thus be seen as an extension of Economides' work of 1984. It will be shown that in most cases this route leads firms to locate at the market quartiles. Hence, the Principle of Almost Intermediate Differentiation is the predominant result.

4.6 Introduction of an outside good[9]

In order to analyze the introduction of a second industry in Hotelling's model of spatial competition we follow Salop's formal setup. The additional industry is purely competitive and produces a homogeneous product. All consumers enjoy the same net surplus, s, if they purchase this commodity. Therefore, any consumer buys the *differentiated* product only if

$$\max_i U(h^*, h_i) \geq s, \tag{4.13}$$

where $U(h^*, h_i)$ is as defined in Section 4.2. Inserting then (4.1) into (4.13) results in

$$\max_i \{v - t|h^* - h_i| - p_i\} \geq 0, \tag{4.14}$$

where $v = u - s$ is called the *reservation* price. Observe that if the full price of the differentiated commodity (that is, mill price plus transportation cost) exceeds the reservation price, the differentiated commodity will not be purchased.

---

[9] This section is based on Hinloopen and Van Marrewijk [1995].

For future analyses it will be convenient the parameterize the model such that

$$l = \alpha\frac{v}{t}.$$ **(4.15)**

The parameter α is thus the size of the market relative to the effective reservation price (v/t), and can be interpreted as an indicator of the equilibrium intensity of rivalry (in a sense supplementary to the Lerner index or the price-cost margin). In particular, the lower is α the fiercer is competition. This can be seen in three different ways.

First, for given values of v and l, two monopolies arise if the cost of transportation t is high, that is, if α is high.[10] Intuitively this is clear. If it is costly for a consumer to switch from one supplier to another, neither firm in the differentiated market experiences fierce competition from the other and both are able to earn monopoly profits. On the other hand, if transportation cost falls (and hence α falls), firms will compete more and more on the basis of the price they quote. In the extreme case of zero transportation cost (that is, $\alpha=0$) the market will yield competitive prices.[11] Second, if the market is large, relative to the effective reservation price (indicating a 'high' α), there is room for two monopolies, while for relatively small markets (or a 'small' α) firms target their sales at almost the same consumers. To put it differently: if the market is large there is a wide dispersion in demand allowing each firm to exclusively serve some part of total demand, while a small market indicates that consumers prefer similar products inducing both firms to compete for the same customers. Third, if the reservation price increases, relative to l and t (that is, α shrinks), demand for the differentiated commodity becomes more and more inelastic. This could be due to an increase in the surplus experienced from the differentiated commodity (u), a decrease in the net surplus associated with the homogeneous good (v), or both. If the reservation price is 'sufficiently high' (or α being 'sufficiently' low) demand is completely inelastic and we are in Hotelling's model, inducing both firms to start a price war. On the other hand, in case of elastic demand (corresponding to a 'low' reservation price or a 'high' α) only a few (local) consumers are willing to purchase the differentiated commodity and no competition is witnessed between the two firms.

In what follows we analyze the market equilibria for different values of α. In particular, we allow α to vary between zero and infinity and determine for this whole range what price firms will quote, in the neighbourhood of $h_a = h_b$, and where they consequently will locate.

4.6.1 Strategic zero conjectural variations

Among other things[12] Hotelling (and most of his disciples) assumed that each firm can alter one of its strategic variables (price and location) on the presumption that its rival's control variables remain

---

[10] Of course, in case of two monopolies some consumers do not buy the differentiated product (except for $\alpha=2$, as will become clear when we proceed). In that sense, both firms are in competition with the outside good.

[11] As will be shown, for values of α below 2/3 we are back in Hotelling's world.

[12] See Phlips and Thisse [1982, p.3-4].

fixed; the so-called Zero Conjectural Variation (ZCV) assumption.[13] As we have seen, this will lead firms to undercut each other and no equilibrium in price and location exists. An obvious way to rule out undercutting behaviour is simply to assume that firms will abstain from it. This has led Eaton (1972, p.269) to introduce a modified ZCV.

Definition 4.1 Modified Zero Conjectural Variation (Eaton)

> Each producer in setting his price and location assumes that the other will remain in the same location, charging the same price, subject to the qualification that the action of one producer does not completely eliminate the other

Observe that the possibility of undercutting is ruled out since this will drive either firm out of business. In effect, Eaton's modified ZCV restricts either firm's second stage action set to the quasi-concave part of the concomitant profit function. In Hotelling's original analysis it means that firm a will never quote a price below $\bar{p}_b - t(l - h_a - h_b)$. And because firm b acts according to the same modified ZCV, firm a's price will never be above $\bar{p}_b + t(l - h_a - h_b)$ (see Figure 4.2). Applying then this modified ZCV to the original analysis of Hotelling immediately restores his Principle of Minimum Differentiation.[14]

Eaton's modified ZCV can be seen as a form of tacit collusion to quote a price above marginal cost. Once both firms are located within the market quartiles they are confronted with a classic prisoner's dilemma: undercutting is a dominant strategy although both firms are better off if they both refrain from doing so. Just assuming then that firms will not follow their dominant strategy is of course one way to get round the prisoner's dilemma, although it is far from elegant. The doubt raised earlier as to firms earning positive profits in a Bertrand-like market is still valid.

A crucial point here is that each firm assumes the other firm not to play its dominant strategy. More satisfactory, we think, is to assume that each firm does not give the other firm the *opportunity* to play its dominant strategy. We call this a Strategic ZCV.

Definition 4.2 Strategic Zero Conjectural Variation

> Each producer in setting his price and location assumes that the other will remain in the same location, charging the same price, subject to the qualification that the action of one producer does not let him be completely eliminated by the other.

Again the possibility of undercutting is ruled out, but now because no firm places itself in a position where it could be undercut. As in the case of modified ZCVs, each player's second stage action set

---

[13] See Martin [1993, p.27-28] for a critical review of ZCVs.

[14] See for example Novshek [1980] who uses Eaton's modified ZCV to proof that on a circle firms will locate symmetrically. Kats [1995] however proves the same result without having to assume that firms modify their ZCV.

is restricted to the quasi-concave part of the profit function. But the restriction is achieved in an opposite way. In terms of Figure 4.2 firm a will never locate such that its price will be above $\bar{p}_b + t(l - h_a - h_b)$. Assuming firm b to follow the same strategy implies that it will never locate such that firm a's price will be below $\bar{p}_b - t(l - h_a - h_b)$. Contrary to endowing firms with a modified ZCV however, in a symmetric equilibrium both firms will locate at the market quartiles, not at the centre.

Applying the concept of strategic ZCV to Economides' analysis of 1986, for example, implies that the locus of optimal locations is also defined for values of γ between 1 and 1.26 (see Figure 4.4). In particular, for this range of γ optimal locations are approximately given by $1.40 - 2.76\gamma + 2.08\gamma^2 - 0.46\gamma^3$ (see Economides [1986, p.70]).

4.6.2 Partial subgame perfect equilibrium

To formally describe the equilibrium of the game analyzed in the next section we follow d'Aspremont et al. [1983] and Friedman [1988]. Denote the strategy set of player i in the second stage by $S_i = \{p_i \mid 0 \le p_i \le v\}$, $v \in \mathbf{R}_+$. Given that fixed and marginal costs are zero, player i's second stage payoff function is

$$\pi_i(p_i, p_j; h_i, h_j) = p_i(h_i, h_j) D_i(p_i, p_j; h_i, h_j), \tag{4.16}$$

where $D_i(p_i, p_j; h_i, h_j)$ is the demand (given locations) of player i. Define the best reply correspondence for player i in the second stage game as

$$P_i(p_j(h_i, h_j); h_i, h_j) = \arg\max_{p_i} \pi_i(p_i, p_j; h_i, h_j), \tag{4.17}$$

whenever it is non-empty. The pair of strategies, $p_i^*(h_i, h_j), p_j^*(h_i, h_j)$, is then a *Nash-Cournot Price Equilibrium* if, and only if

$$p_i^*(h_i, h_j) \in P_i(p_j^*(h_i, h_j); h_i, h_j). \tag{4.18}$$

Moving on to the first stage, let $P^*(h_i, h_j)$ be the set of Nash-Cournot Price Equilibria and let Ξ be the subset of pairs (h_i, h_j) for which $P^*(h_a, h_b)$ is non-empty. For any pair $(h_i, h_j) \in \Xi$ we then define

$$\pi_i^*(h_i, h_j) \equiv \pi_i(p_i^*(h_i, h_j), p_j^*(h_i, h_j), h_i, h_j), \tag{4.19}$$

that is, player i's first stage payoff function given that both adhere to their best responses in the second stage game. Hence, player i's best reply correspondence in the first stage game is defined as

$$H_i(h_j) = arg\max_{h_i} \pi_i^*(h_i,h_j), \tag{4.20}$$

whenever it is non-empty. The pair of strategies $(h_i^*,h_j^*) \in \Xi$ is then a *Nash-Cournot Location Equilibrium* if, and only if

$$h_i^* \in H_i(h_j^*). \tag{4.21}$$

For simplicity we only consider neighbourhoods of $h_i = h_j$ when deriving Nash-Cournot equilibrium prices, and thus restrict attention to *local symmetric* Nash-Cournot location equilibria. In particular, given $\varepsilon > 0$, define $B(h_i;\varepsilon)$ as the closed ε-ball around h_i. For all $(h_i,h_j) \in \Xi$ define player i's local best reply correspondence as

$$H_i(h_j;\varepsilon) = arg\max_{h_i \in B(h_i;\varepsilon)} \pi_i^*(h_i,h_j), \tag{4.22}$$

whenever it is non-empty. The pair of strategies $(h_i^*,h_j^*) \in \Xi$ is then a *Local Symmetric Nash-Cournot Location Equilibrium* if, and only if, $\exists \varepsilon > 0$ such that

$$h_i^* \in H_i(h_j^*;\varepsilon). \tag{4.23}$$

The equilibrium of the whole game is now implicitly defined by (4.18) and (4.23). This could be called a Local Symmetric Nash-Cournot Price-Location Equilibrium.

Because we only consider market configurations in the neighbourhood of $h_a = h_b$, we compute what Friedman [1988] labels partially subgame-perfect equilibria. The equilibrium strategy is an equilibrium for some but not all subgames. He hastens to add however that (p.615)

> It is worth emphasizing that partial subgame perfection does not deserve to be regarded as a new equilibrium concept. It is merely a convenient label to attach to a pure-strategy, noncooperative equilibrium that induces a noncooperative equilibrium on certain, but not all, subgames.
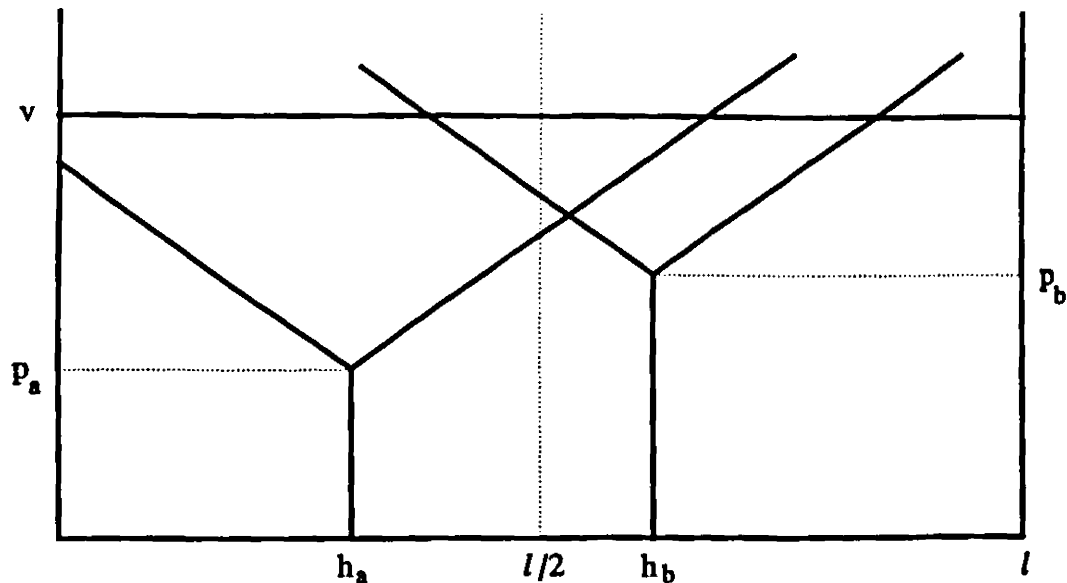
Figure 4.6 Non local symmetric locations

With respect to a subgame-perfect equilibrium it thus seems that we have made two restrictions: (i) we only consider symmetric locations and (ii) we confine ourselves to examining local symmetric locations. The first of these implies that we only consider parts of the two-stage game. A situation as depicted in Figure 4.6 for example will not be analyzed. But if we do not restrict ourselves in this way, the number of cases to be considered in the next section will be far too many. The second constraint on the concept of subgame perfectness is however not a genuine restriction. Clearly the set of symmetric Nash-Cournot price-location equilibria is a subset of the set of local symmetric Nash-Cournot price-location equilibria. Since in all cases considered in the next section, in which the two firms actually compete with each other, there exists only one local symmetric Nash-Cournot price-location equilibrium, computing the partial subgame perfect equilibrium, as implied by (4.18) and (4.23), identifies the unique symmetric Nash-Cournot price-location equilibrium. Moreover, if the market is large enough to sustain two monopolies, our procedure identifies all symmetric Nash-Cournot price-location equilibria.

To summarize, a local symmetric Nash-Cournot location equilibrium, (h_i^*, h_j^*), is reached if either

$$(i) \quad \frac{\partial \pi_i^*(h_i, h_j)}{\partial h_i} = 0,$$

$$(ii) \quad \frac{\partial \pi_i^*(h_i, h_j)}{\partial h_i} = \begin{cases} > 0, & h_i < h_i^* \\ < 0, & h_i > h_i^* \end{cases},$$

(4.24)

$$(iii) \quad \frac{\partial \pi_i^*(h_i, h_j)}{\partial h_i} = \begin{cases} > 0, & h_i < h_i^* \\ N.D., & h_i > h_i^* \end{cases}.$$

The first case corresponds to the first order condition of any maximization problem. In the second situation, the neighbourhood of the peak of the profit-curve resembles something like a tent with a (sharp) ridge. This clearly indicates a maximum at (h_i^*, h_j^*). Both players explicitly follow their strategic ZCV in the third case; driven by profit increasing incentives they move towards each other but never so much that the other player can drive them out of business.

Having defined the equilibrium concept to be used in the sequel of this chapter, we can now address the question as to what the market equilibrium will be in Hotelling's spatial model when an outside good is introduced.

4.6.3 Two monopolies ($2 \leq \alpha < \infty$)

We start our analysis by investigating a market that is large relative to the effective reservation price v/t. In fact, the market will be so large that in equilibrium the two firms will not compete with each other and will earn (maximum) monopoly profits.

Minding Your Own Business

For relatively small values of h_i, the firms are so far apart from each other that they form local monopolies. If firm i is located at h_i its profits are given by (see Figure 4.7)
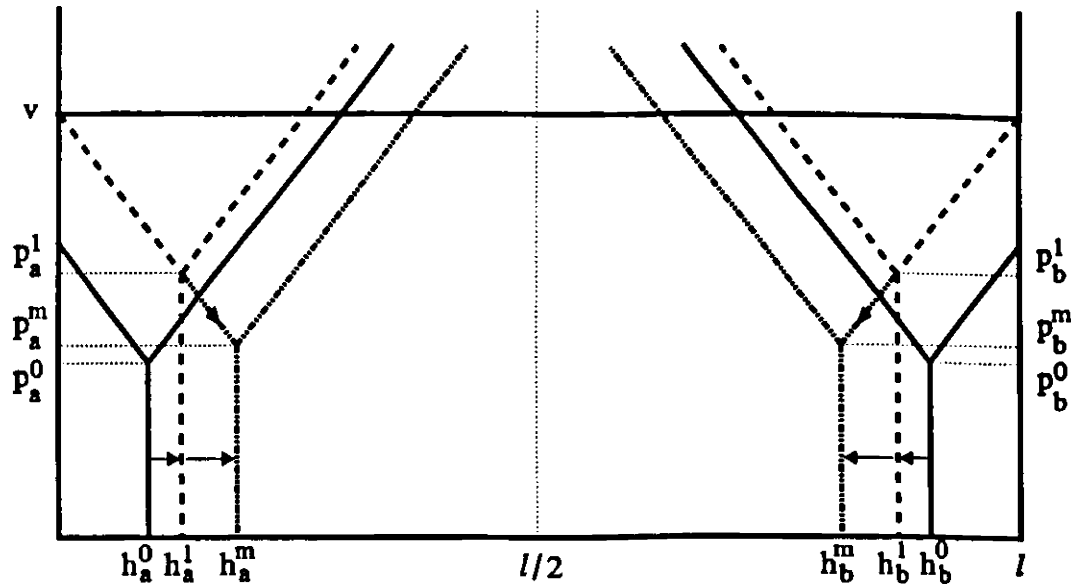
Figure 4.7 Minding your own business

$$\pi_i(p_i, p_j; h_i, h_j) = \begin{cases} p_i\left(h_i + \dfrac{v - p_i}{t}\right), & 0 \leq p_i < \max\{0, v - th_i\}, \\[3mm] 2p_i h_i, & p_i = \max\{0, v - th_i\}, \\[3mm] 2p_i\left(\dfrac{v - p_i}{t}\right), & \max\{0, v - th_i\} < p_i \leq v. \end{cases} \tag{4.25}$$

The first segment of (4.25) corresponds to the case in which the consumer located at the border of the market is confronted with a price below its reservation price, while the location of the consumer who is indifferent between buying or not buying the differentiated product (x, being in between the two firm's locations) is implicitly defined by $p_i + tx = v$. Firm i's total demand in this case thus equals $h_i + (v - p_i)/t$. In Figure 4.7 this situation is envisaged by prices p_a^0 and p_b^0, and locations h_a^0 and h_b^0. If the consumer at the market boundary is indifferent whether or not to purchase the differentiated commodity, the price quoted equals $v - th_i$. Demand in this case equals $2h_i$ (prices p_a^1 and p_b^1, and locations h_a^1 and h_b^1 in Figure 4.7). Finally, if price and location are such that either firm does not (or just) serve(s) its complete hinterland, firm i's demand is given by $2(v - p_i)/t$. In terms of Figure 4.7 this corresponds to prices p_a^m and p_b^m at locations h_a^m and h_b^m respectively.

Profit maximization is now straightforward and leads to Nash-Cournot equilibrium prices

$$
p_i^*(h_i,h_j) = \begin{cases} \dfrac{1}{2}(v+th_i), & 0 \le h_i \le \dfrac{v}{3t}, \\[2ex] v-th_i, & \dfrac{v}{3t} \le h_i \le \dfrac{v}{2t}, \\[2ex] \dfrac{v}{2}, & \dfrac{v}{2t} \le h_i \le \dfrac{l}{2} - \dfrac{v}{2t}. \end{cases}
\tag{4.26}
$$

The intervals for h_i over which $p_i^*(h_i,h_j)$ is defined follow from the partition of $\pi_i(p_i,p_j;h_i,h_j)$ according to the value of p_i (see (4.25)). They can also be retrieved by observing that the three parts of which the equilibrium price is made up respectively corresponds to h_i being less than, equal, or greater than $(v-p_i)/t$. The upper bound in terms of h_i over which the equilibrium price is defined, $l/2 - v/2t$, will be explained below.

First stage profits are now given by

$$
\pi_i^*(h_i,h_j) = \begin{cases} \dfrac{1}{4t}(v+th_i)^2, & 0 \le h_i < \dfrac{v}{3t}, \\[2ex] 2h_i(v-th_i), & \dfrac{v}{3t} \le h_i \le \dfrac{v}{2t}, \\[2ex] \dfrac{v^2}{2t}, & \dfrac{v}{2t} < h_i < \dfrac{l}{2} - \dfrac{v}{2t}. \end{cases}
\tag{4.27}
$$

From these it is readily derived that $\partial \pi_i^*(h_i,h_j)/\partial h_i$, $i,j=a,b$, $i \ne j$, is positive for $h_i < v/2t$, implying that either firm's profits are maximized if its location is at least a distance $v/2t$ away from the market boundary. The total length of the market covered by one firm is then v/t and the price charged equals $v/2$. Both firms can thus earn monopoly profits if the market is large enough relative to the effective reservation price, that is *iff* $l \ge 2v/t$. This is equivalent to $\alpha \ge 2$, the case under investigation in this section. Finally, notice that if both firms move towards the middle of the market, assuming that they are located symmetrically around $l/2$, they start encroaching on each others' customers once h^* ($=h_a^*=h_b^*$) exceeds $l/2 - v/2t$. Examination of what happens then is in the next subsection.

The Economides Case

What happens if firms are located such that they start encroaching on each other's customers, while maintaining the assumption that some customers at the corners of the market are not served, has been

analyzed by Economides [1984]. He derives the following second stage profit function for firm i

$$\pi_i(p_i,p_j;h_i,h_j) = \begin{cases} 2p_i\left(\dfrac{v-p_i}{t}\right), & 0\le p_i\le p_j-t(l-h_i-h_j), \\[2ex] p_i\left(\dfrac{2v-3p_i+p_j}{2t}+\dfrac{l-h_i-h_j}{2}\right), & \\[2ex] & p_j-t(l-h_i-h_j)\le p_i\le 2v-p_j-t(l-h_i-h_j), \\[2ex] 2p_i\left(\dfrac{v-p_i}{t}\right), & 2v-p_j-t(l-h_i-h_j)\le p_i\le v. \end{cases} \quad (4.28)$$

The first segment of this profit function is determined by the possibility of undercutting one's competitor. If $p_i\le p_j-t(l-h_i-h_j)$ all consumers who buy the differentiated commodity will do so from firm i. The middle part of (4.28) corresponds to the case in which the reservation price of the marginal consumer (located between the two firms) is higher than the mill price plus transportation cost he faces from either firm. Firm i's demand is then the sum of $(v-p_i)/t$ and $[p_j-p_i+t(l-h_i-h_j)]/2t$. Finally, if the marginal consumer is indifferent between buying from either firm or not to buy at all, both firms just cover the market between them. In particular, $(v-p_i)/t+(v-p_j)/t\le l-h_i-h_j$ or $2v-p_j-t(l-h_i-h_j)\le p_i$.

For this setting Economides [1984] identifies three types of Nash-Cournot price-equilibria (see Figure 4.8). First there are local monopolistic equilibria. As long as firms are not in competition for the same customers they can both earn monopoly profits. In particular, for $p_i\in(2v-p_j-t(l-h_i,h_j),v]$ firm i's profits equal $2p_i(v-p_i)/t$ and are maximized for $p_i=v/2$. Of course, this equilibrium price corresponds to the third segment of $p_i^*(h_i,h_j)$, as derived in the previous subsection (see equation (4.26)). Notice that $p_i^*\ge 2v-p_j^*-t(l-h_i-h_j)$ requires that $v/t\le(l-h_i-h_j)$ if $p_i^*=p_j^*=v/2$.

Maximizing the second part of (4.28) with respect to p_i and solving for the symmetric solution leads to $p_i^*(h_i,h_j)=p_j^*(h_i,h_j)=[2v+t(l-h_i-h_j)]/5$. Economides [1984] calls this the competitive Nash-Cournot price equilibrium. Competitive profits are then equal to $3([2v+t(l-h_i-h_j)]/5)^2/2t$. To examine if the competitive price is indeed a Nash-Cournot equilibrium price we first have to determine when undercutting becomes profitable. The price firm i has to quote to serve the whole market, $p_i''(h_i,h_j)$, follows from $p_i''(h_i,h_j)+t(l-h_i-h_j)=p_j^*(h_i,h_j)$. The undercutting price equals $p_i''(h_i,h_j)=2[v-2t(l-h_i-h_j)]/5$. Profits are then given by $2p_i''(h_i,h_j)[v-p_i''(h_i,h_j)]/t$, and are less than competitive profits *iff* $(l-h_i-h_j)<6v/t(7+5\sqrt{10})$. Further, as derived above, the part of firm i's profit function under consideration requires that its price is in a certain interval. In particular, it must set its price such that $p_i\in[p_j-t(l-h_i-h_j),2v-p_j-t(l-h_i-h_j)]$. The two boundaries determining this price-interval give rise to $|p_i-p_j|\le t(l-h_i-h_j)$ and $p_i+p_j\le 2v-t(l-h_i-h_j)$. For the symmetric competitive prices the first of these is obviously fulfilled. The second condition implies that $6v/7t\le(l-h_i-h_j)$.
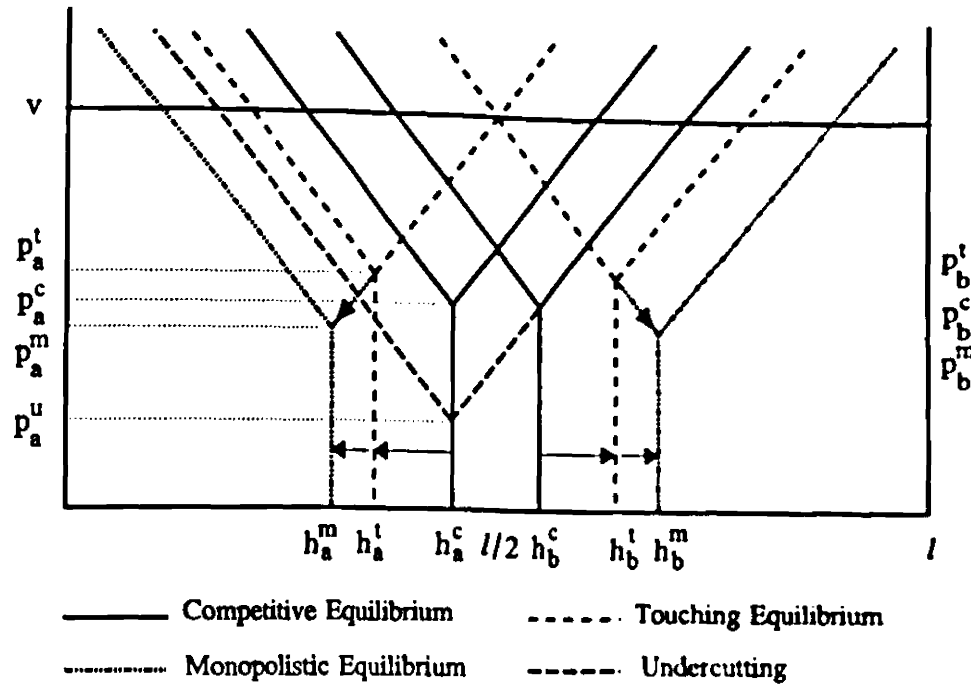
Figure 4.8 The Economides case

If both firms together just supply all consumers located between the two firms, a Nash-Cournot price equilibrium exists, which Economides [1984] calls a touching equilibrium. As explained above, in this situation the marginal consumer is indifferent between buying the differentiated commodity from either firm or not buying it at all. It is at this consumer's location where both firms' market regions touch. In this case firm i's price is implicitly determined by $p_i^*(h_i,h_j)+t(l-h_i-h_j)/2=v$, while its market share is $2[v-p_i^*(h_i,h_j)]/t$. For what locations the proposed price is indeed a Nash-Cournot equilibrium price is determined by profit considerations. In particular, touching equilibrium profits equal $t[v/t-(l-h_i-h_j)/2](l-h_i-h_j)$. These exceed competitive profits whenever $(l-h_i-h_j)>6v/7t$. On the other hand, the distance between both firms, $(l-h_i-h_j)$, must be at least v/t for them to be able to earn monopoly profits.

These considerations give rise to the following Nash-Cournot equilibrium price

$$
p_i^*(h_i,h_j) = \begin{cases} \dfrac{2v+t(l-h_i-h_j)}{5}, & \dfrac{6v}{t(7+5\sqrt{10})}<(l-h_i-h_j)<\dfrac{6v}{7t}, \\[3mm] v-\dfrac{t}{2}(l-h_i-h_j), & \dfrac{6v}{7t}<(l-h_i-h_j)<\dfrac{v}{t}, \\[3mm] \dfrac{v}{2}, & \dfrac{v}{t}<(l-h_i-h_j)<l-\dfrac{v}{t}. \end{cases}
$$

(4.29)

Notice that for $(l-h_i-h_j)>l-v/t$ two monopolies can no longer be sustained and the analysis of the previous subsection applies. First stage profits are now given by

$$
\pi_i^*(h_i,h_j) = \begin{cases} \dfrac{3}{2t}\left(\dfrac{2v+t(l-h_i-h_j)}{5}\right)^2, & \dfrac{6v}{t(7+5\sqrt{10})}<(l-h_i-h_j)<\dfrac{6v}{7t}, \\[4ex] [v-\dfrac{t}{2}(l-h_i-h_j)](l-h_i-h_j), & \dfrac{6v}{7t}<(l-h_i-h_j)<\dfrac{v}{t}, \\[4ex] \dfrac{v^2}{2t}, & \dfrac{v}{t}<(l-h_i-h_j)<l-\dfrac{v}{t}. \end{cases}
\tag{4.30}
$$

It is straightforward to show that $\partial\pi_i^*(h_i,h_j)/\partial h_i$, is negative for $(l-h_i-h_j)<v/t$. That is, neither firm wants to locate too close to the other in order to be able to earn monopoly profits.

For completeness we will give the Nash-Cournot equilibrium prices and locations for the values of α under consideration

Lemma 4.1

(i) For $2\leq\alpha<\infty$ the symmetric Nash-Cournot equilibrium price in the neighbourhood of $h_a=h_b$ equals

$$
p_i^*(h_j,h_j) = \begin{cases} \dfrac{1}{2}(v+th_i), & 0\leq h_i\leq\dfrac{v}{3t}, \\[3ex] v-th_i, & \dfrac{v}{3t}\leq h_i\leq\dfrac{v}{2t}, \\[3ex] \dfrac{v}{2}, & \dfrac{v}{t}\leq(l-h_i-h_j)\leq l-\dfrac{v}{t}, \\[3ex] v-\dfrac{t}{2}(l-h_i-h_j), & \dfrac{6v}{7t}\leq(l-h_i-h_j)\leq\dfrac{v}{t}, \\[3ex] \dfrac{2v+t(l-h_i-h_j)}{5}, & \dfrac{6v}{t(7+5\sqrt{10})}<(l-h_i-h_j)\leq\dfrac{6v}{7t}. \end{cases}
$$

(ii) For $2\leq\alpha<\infty$ the symmetric Nash-Cournot equilibrium locations , h_a^* and h_b^*, are such that

$$
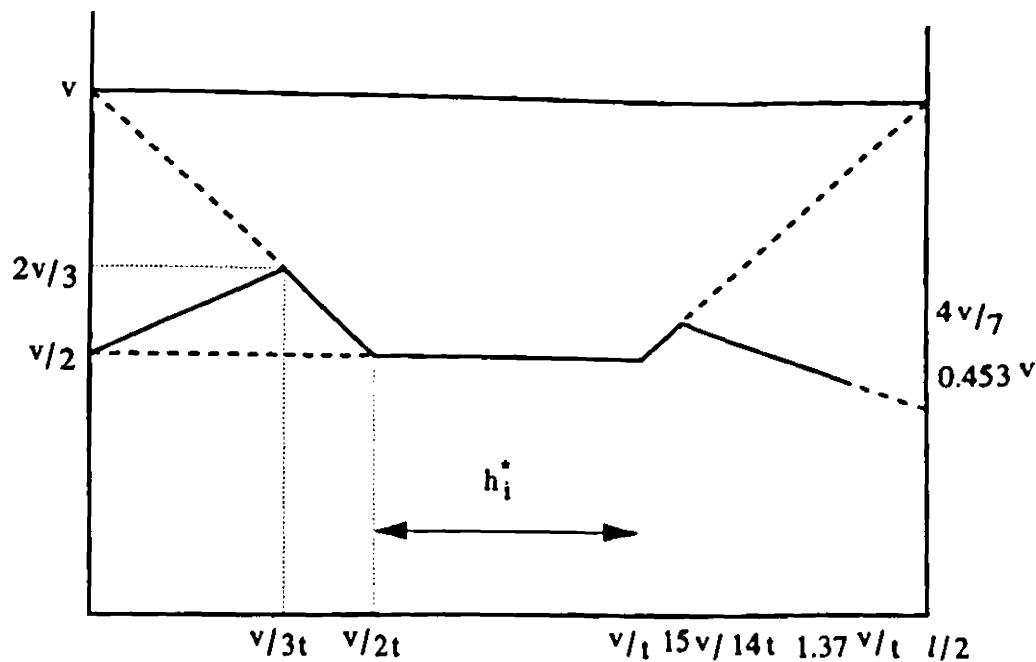\frac{v}{2}\leq(l-h_i^*-h_j^*)<l-\frac{v}{t}.
$$

Figure 4.9 Symmetric Nash-Cournot pricing strategy for $\alpha = 3$

Note that in case of two local monopolies the locations need not to be symmetric. However, in deriving the equilibrium price we only considered neighbourhoods of $h_a = h_b$. The validity of the second part of Lemma 4.1 is therefore restricted to symmetric Nash-Cournot location equilibria.

Figure 4.9 illustrates the complete analysis of this section for the case $\alpha \approx 3$. It depicts firm i's Nash-Cournot price equilibrium strategy, p_i^*, as a function of its location ($0 \le h_i < l/2$) for symmetric locations of the two firms ($h_a = h_b$), if these exist. Some illustrative prices are also drawn. The picture of the market could be completed with regard to the other firm's Nash-Cournot price equilibrium strategy by putting a mirror at $l/2$. The continuum of symmetric Nash-Cournot price-location equilibria are at $h_i^* \in [v/2t, l/2 - v/2t]$, with concomitant equilibrium price $v/2$. For Figure 4.4, where α equals 3, this translates to $h_i^* \in [v/2t, v/t]$. Finally, pure strategy Nash-Cournot equilibrium prices are defined up to $h_i = l/2 - 3v/t(7 + 5\sqrt{10}) \approx 1.368 v/t$, at which location they are $0.453 v$.

4.6.4 Being aware of each other ($4/3 \le \alpha \le 2$)

We proceed our discussion with analyzing what happens if α falls below the range considered in Section 4.6.3. An inspection of Figure 4.9 shows that as α falls the range over which h_i^* is defined starts to shrink. This may suggest that once α is below 2, this range disappears; that is, an equilibrium in location no longer exists. And indeed, as mentioned in Section 4.4.2, Economides [1984] conjectures that for a Nash equilibrium of the varieties game to exist, it is essential that some

customers at the edges of the market are not served. In what follows however, we show that h_i^* is also defined when both firms together cover the entire market.

If both firms are located near the edges of the market, the reasoning of the previous subsection still applies. Firms then want to move towards each other in order to gain market shares. But because the market is now small relative to the effective reservation price, there is no room for two monopolies. Before either firm can charge the monopoly price, $v/2$, and locate accordingly, Economides' touching equilibrium is reached (see Figure 4.10a). This always happens when firms are located at the market quartiles.[15]

On the other hand, if the distance between the two firms is small, we can still apply Economides' analysis. As is shown in the previous subsection, firms then have an incentive to move away from each other. However, the transition of Economides' touching equilibrium to the monopolistic equilibrium will be incomplete. Before either firm can act as a local monopolist, the whole market is served (see Figure 4.10b). This happens exactly at the market quartiles.

Once both firms are at the market quartiles, neither wants to relocate. Observe that this is according to the second 'definition' of a local symmetric Nash-Cournot location equilibrium (see equation (4.24)).

Because the market considered is now small relative to the effective reservation price, there are two refinements to be made with respect to Economides' analysis.

First, he implicitly assumes that when a firm undercuts its opponent, its market share will be $2(v-p_i^u)/t$, with p_i^u as undercutting price. But if the market is relatively small, it can be that $h_i < (v-p_i^u)/t$ (see Figure 4.10b, where $h_u^c < (v-p_u^u)/t$). The undercutting rule then changes; the mathematical details are delegated to Appendix 4A.

Second, when moving from Economides' competitive equilibrium to that at the market quartiles, Economides' touching equilibrium might not be reached (see Figure 4.11). For this equilibrium to be feasible it must be that $(v-p_i^c)/t \leq h_i$, where p_i^c is the competitive price charged by firm i, as defined in the previous sub section (see Lemma 4.1). This condition, together with the observation that locations are within the market quartiles, translates into $\alpha \geq 12/7$. For smaller values of α the consumers at the market boundaries will always be served *before* the market quartiles are reached. This equilibrium, called a competitive equilibrium with full supply, will be examined in more detail in the next subsection. There it will also be shown that for this situation with $\alpha \in (4/3, 12/7)$ both firms have an incentive to move away from each other. Hence, their location at the market quartiles.

Lemma 4.2 summarizes the discussion of this subsection.

---

[15] In the next subsections we will consider cases in which firms cover the entire market even before they are located at the market quartiles.
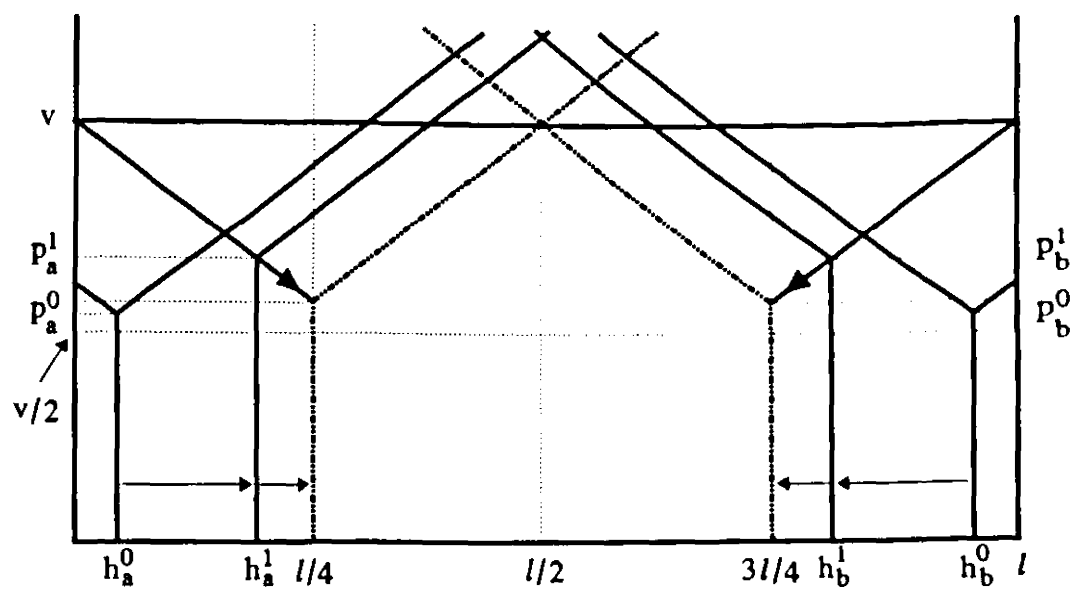
Figure 4.10a Movement towards market quartiles from hinterlands
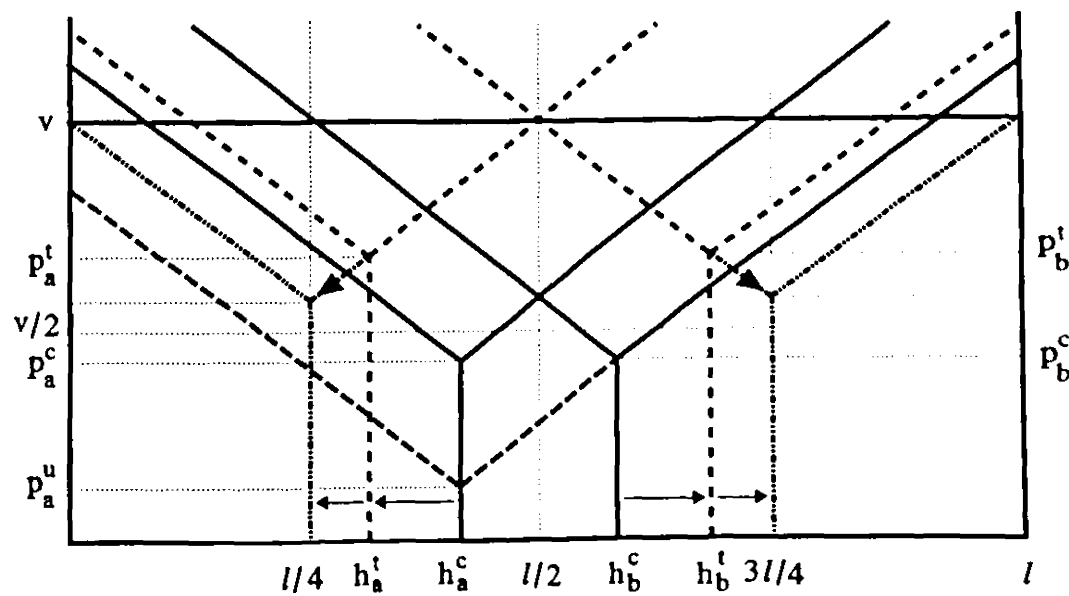
Figure 4.10b Movement towards market quartiles from market centre

Lemma 4.2

(i) For $4/3 \leq \alpha \leq 2$ the symmetric Nash-Cournot equilibrium price in the neighbourhood of $h_a = h_b$ equals

$$
p_i^*(h_i, h_j) = \begin{cases}
\dfrac{1}{2}(v + th_i), & 0 \leq h_i \leq \dfrac{v}{3t}, \\[2ex]
v - th_i, & \dfrac{v}{3t} \leq h_i \leq \dfrac{l}{4}, \\[3ex]
\end{cases} \qquad \dfrac{4}{3} \leq \alpha \leq 2.
$$

$$
\begin{cases}
v - \dfrac{t}{2}(l - h_i - h_j), & \dfrac{6v}{7t} \leq (l - h_i - h_j) \leq \dfrac{l}{2}, \\[2ex]
\dfrac{2v + t(l - h_i - h_j)}{5}, & \dfrac{6v}{t(7 + 5\sqrt{10})} < (l - h_i - h_j) \leq \dfrac{6v}{7t},
\end{cases} \qquad \dfrac{6(4 + \sqrt{10})}{7 + 5\sqrt{10}} \leq \alpha \leq 2.
$$

$$
\begin{cases}
v - \dfrac{t}{2}(l - h_i - h_j), & \dfrac{6v}{7t} \leq (l - h_i - h_j) \leq \dfrac{l}{2}, \\[3ex]
\dfrac{2v + t(l - h_i - h_j)}{5}, & \dfrac{2}{7}\left[\sqrt{\left(\dfrac{v}{t} + 2h_i\right)^2 + 7\dfrac{v}{t}h_i} - \dfrac{v}{t} - 2h_i\right] \leq (l - h_i - h_j) \leq \dfrac{6v}{7t},
\end{cases} \qquad \dfrac{12}{7} \leq \alpha \leq \dfrac{6(4 + \sqrt{10})}{7 + 5\sqrt{10}}.
$$

$$
\begin{cases}
v - th_i, & \dfrac{l}{2} - 2h_i + \dfrac{l(v - tl + th_i)}{v - th_i} \leq (l - h_i - h_j) \leq \dfrac{l}{2}, \\[3ex]
\dfrac{2v + t(l - h_i - h_j)}{5}, & \dfrac{2}{7}\left[\sqrt{\left(\dfrac{v}{t} + 2h_i\right)^2 + 7\dfrac{v}{t}h_i} - \dfrac{v}{t} - 2h_i\right] \leq (l - h_i - h_j) \leq \dfrac{l}{2} - 2h_i + \dfrac{l(v - tl + th_i)}{v - th_i}.
\end{cases} \qquad \dfrac{4}{3} \leq \alpha \leq \dfrac{12}{7}.
$$

(ii) For $4/3 \leq \alpha \leq 2$ the symmetric Nash-Cournot equilibrium locations are

$$
h_i^* = \frac{l}{4}.
$$
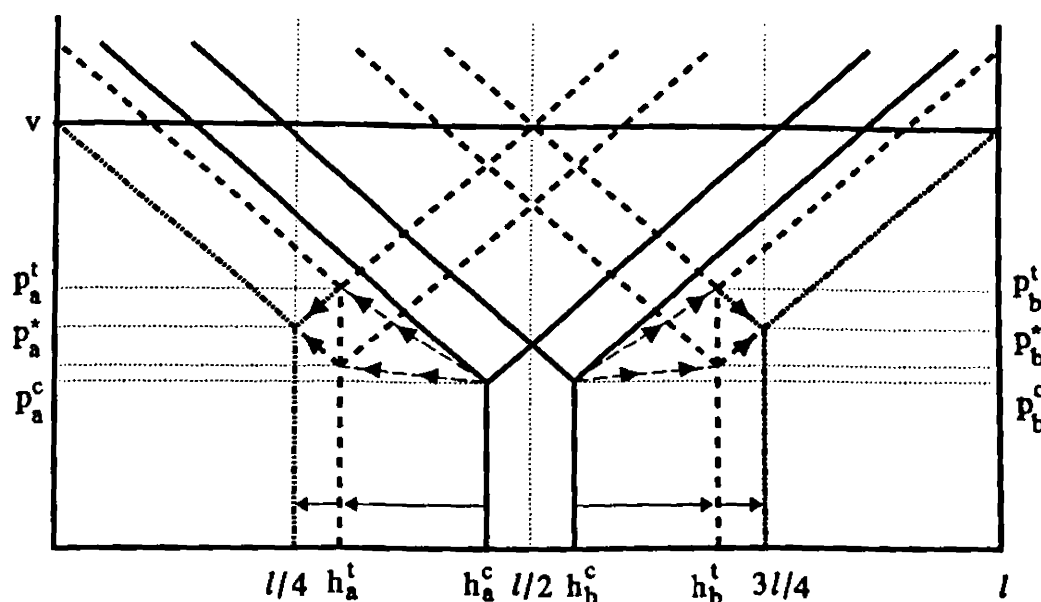
Figure 4.11 Moving from Economides' competitive equilibrium to the market quartiles, through (i) his touching equilibrium ($\alpha > 12/7$) or, (ii) a competitive equilibrium with full supply ($\alpha < 12/7$)

So far we have seen that introduction of a second industry leads firms to form local monopolies or to locate at the market quartiles if they are in competition with each other. In either case they do not locate back to back at the market centre. Rather, they produce differentiated products, thereby invalidating the Principle of Minimum Differentiation. Moreover, an equilibrium in the varieties stage was shown to exist, even if both firms together supply the whole market. We proceed with examining what happens if competition becomes even more fierce, that is, if α drops below $4/3$.

4.6.5 Getting close ($8/7 \leq \alpha \leq 4/3$)

For the range of α under consideration in this subsection, it is still possible for firms not to compete when located 'far enough' from each other (in terms of Figure 4.10a this means that the situation characterized by prices p_a^0, p_b^0, and locations h_a^0, h_b^0 is still feasible). Firms then have an incentive to move toward each other, as shown in Section 4.6.3. But in contrast to the analysis of the previous subsection, for α below $4/3$ the first conflict between the two firms always arises *before* they reach the market quartiles. This situation is the analogue of Economides' touching equilibrium when consumers at the corners of the market are served rather than not served.

A Touching Equilibrium With Full Supply

If both firms are located outside the market quartiles they together supply the entire market
$(v-p_i)/t+(v-p_j)/t \geq 2l-h_i-h_j$. This is equivalent to $l/2 \geq h_i+h_j \geq 2l-2v/t$, or $\alpha \leq 4/3$, if we recall that
in absence of competition each firm's price equals $(v+th_i)/2$. That is, only for $\alpha \leq 4/3$ is it possible
that firms compete while located outside the market quartiles. In this case, firms' market regions touch
in the market centre (for symmetric locations) and all consumers buy the differentiated product (see
Figure 4.12). Firm i's second stage price is then given by

$$p_i^*(h_i,h_j) = v - \frac{(l-h_i-h_j)}{2},$$

$$(4.31)$$

with concomitant (touching) profits

$$\pi_i^*(h_i,h_j) = \frac{t}{2}\left(\frac{v}{t}-\frac{l-h_i-h_j}{2}\right)(l+h_i-h_j).$$

$$(4.32)$$

Considering now the first stage, observe that $\partial \pi_i^*(h_i,h_j)/\partial h_i$ is positive. That is, firms want to move
toward each other (see Figure 4.12). Once both firms reach the market quartiles, a different equilibrium
arises which will be analyzed below.

But before we proceed we have to examine under which conditions the touching equilibrium
prices are indeed Nash-Cournot prices in the neighbourhood of $h_i=h_j$, because if not, nothing can be
said about the tendency to agglomerate. First, note that undercutting requires a price equal
$p_i^u(h_i,h_j)=v-3t(l-h_i-h_j)/2$. For symmetric locations, profits when undercutting, $p_i^u(h_i,h_j)$, are
below touching profits as long as $h_i=h_j \leq l/2-v/5t$. Now observe that $l/4 \leq l/2-v/5t$ whenever
$\alpha \geq 4/5$. But because $h_i=h_j \leq l/4$ it is true that $h_i=h_j \leq l/2-v/5t$ for $\alpha \geq 4/5$. Clearly this rules out
the threat of undercutting for the range of α under consideration.[16] Second, are the proposed
equilibrium prices stable? That is, given the touching equilibrium price quoted by one firm, does the
other have an incentive to change its price by a small fraction? If firm i increases its price by a
fraction ε, its profits are

$$\pi_i^*(h_i,h_j) = \frac{t}{2}\left(\frac{v}{t}-\frac{l-h_i-h_j}{2}+\frac{\varepsilon}{t}\right)\left(l+h_i-h_j-\frac{2\varepsilon}{t}\right).$$

$$(4.33)$$

given that firm j's price remains fixed (see Figure 4.13). These are less than touching profits.

---

[16] Moreover, if $0 \leq h_i+h_j \leq l(3\alpha-2)/3\alpha$, undercutting requires a negative price.
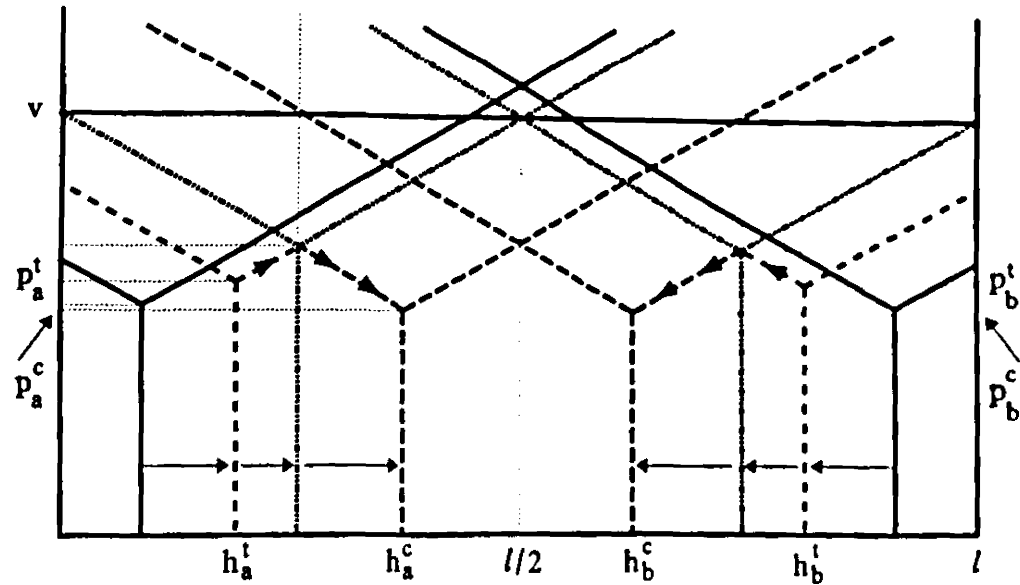
Figure 4.12 **(i) A touching equilibrium with full supply** (h_i^t, p_i^t) **and (ii) A competitive equilibrium with full supply** (h_i^c, p_i^c)

$v/t \geq l - h_j$. Because both firms are located outside the market quartiles this condition translates into $\alpha \leq 4/3$. Indeed, if $\alpha \geq 4/3$ firms have an incentive to raise their price and the situation analyzed in Section 4.6.4 arises. On the other hand, if firm i slightly lowers its price while that of firm j does not change, its profits are

$$\pi_i^-(h_i, h_j) = \frac{t}{2}\left(\frac{v}{t} - \frac{l - h_i - h_j}{2} - \frac{\varepsilon}{t}\right)\left(l + h_i - h_j + \frac{\varepsilon}{t}\right) \tag{4.34}$$

(see Figure 4.13). Touching profits exceed these *iff* $3l + h_i - 3h_j \geq 2v/t$. For symmetric locations this requires that $h_i = h_j \leq 3l/2 - v/t$, or $\alpha \geq 4/5$, if we recall that both firms are located outside the market quartiles. Observe that if it is profitable to lower prices for any pair of (symmetric) locations, the situation analyzed by Hotelling emerges.[17]

In short, the Nash-Cournot equilibrium prices for the touching equilibrium in which all consumers are served, equal $p_i^*(h_i, h_j) = v - t(l - h_i - h_j)/2$. Concomitant profits are $\pi_i^*(h_i, h_j) = [v - t(l - h_i - h_j)/2][h_i + (l - h_i - h_j)/2]$, implying that both firms want to move toward the market centre.

---

[17] This situation occurs when $\alpha \leq 2/3$, and will be analyzed in Section 4.6.7.
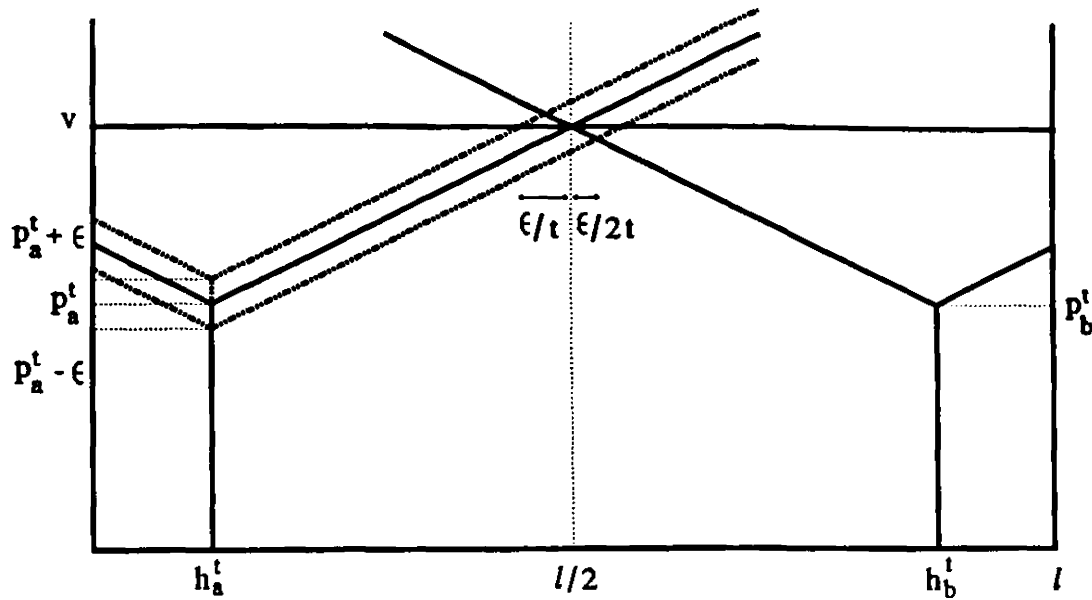
Figure 4.13 Price stability in a touching equilibrium with full supply

A Competitive Equilibrium With Full Supply

Once firms reach the market quartiles, Nash-Cournot prices are given by

$$p_i^*(h_i,h_j) = v - th_i,$$ (4.35)

(recall from Section 4.6.4 that because $\alpha \leq 12/7$ Economides' touching equilibrium will not be reached). Firm i's second stage profits are then equal to

$$\pi_i^*(h_i,h_j) = (v - th_i)\left(\frac{l}{2} + h_i - h_j\right).$$ (4.36)

We will call these competitive profits with full supply (see Figure 4.12). Reaction functions for the first stage follow from $\partial \pi_i^*(h_i,h_j)/\partial h_i = 0$, and are given by[18]

---

[18] The equilibrium of the first stage is stable whenever the reaction functions cross correctly, that is whenever $|\partial h_i(h_j)/\partial h_j| < 1$, which is the case here.

$$h_i(h_j) = \frac{h_j}{2} - \frac{l}{4} + \frac{v}{2t}. \tag{4.37}$$

Local symmetric Nash-Cournot locations are then readily derived

$$h_i^* = \frac{v}{t} - \frac{l}{2}. \tag{4.38}$$

Observe that these equilibrium locations follow from the first 'definition' of a local symmetric Nash-Cournot location equilibrium (see equation 4.24).

Again we have to check whether or not the proposed equilibrium prices are indeed Nash-Cournot prices. The price firm i has to quote to undercut firm j and hence to supply the whole market, $p_i^u(h_i,h_j)$, follows from $p_i^u(h_i,h_j) + t(l - h_i - h_j) = v - th_i$, and equals $p_i^u(h_i,h_j) = v - tl + th_i$. Its profits are then equal to $\pi^u(h_i,h_j) = [v - tl + th_i]l$. Undercutting will not be profitable as long as these profits do not exceed competitive profits, that is if $[v/t - h_i][3l + 2(h_i - h_j)] \geq 2l(2v/t - l)$. For symmetric locations this boils down to

$$h_i^* \leq \frac{2l}{3} - \frac{v}{3t}. \tag{4.39}$$

Given that $h^* = v/t - l/2$ it must be that $\alpha \geq 8/7$ for undercutting not to be profitable. That is, for $4/3 \leq \alpha \leq 8/7$, the expression in (4.39) determines where firms will locate.[19] Further, as before, we have to consider the stability of the proposed equilibrium prices. Does either firm want to slightly change its price, given that its competitor sticks to the equilibrium price? Let

$$\pi_i^+(h_i,h_j) = (v - th_i + \varepsilon)\left(\frac{l}{2} + h_i - h_j - \frac{3\varepsilon}{2t}\right) \tag{4.40}$$

be the profits firm i receives when it raises its price by an amount ε. It will do so whenever these profits exceed competitive profits, that is if $2(v/t - h_i) \leq (l/2 + h_i - h_j)$. It then follows that for symmetric locations, the proposed equilibrium prices are valid for

$$h_i^* \leq \frac{v}{t} - \frac{l}{3}. \tag{4.41}$$

---

[19] Remember that for the range of α under consideration firms will never locate outside the market quartiles, as shown in the discussion of the touching equilibrium with full supply.

a condition which is clearly satisfied.[20] Lowering the price by a small amount results in profits

$$\pi_i^-(h_i,h_j) = (v-th_i-\varepsilon)\left(\frac{l}{2}+h_i-h_j+\frac{\varepsilon}{2t}\right).$$

(4.42)

These are below competitive profits if $3h_i-2h_j \geq v/t-l$. In case of symmetric locations this condition implies that

$$h_i^* \geq \frac{v}{t}-l,$$

(4.43)

which holds for the derived equilibrium locations. Observe that Hotelling's original analysis applies if it is profitable to lower the price for any pair of (symmetric) locations, a situation which arises *iff* $\alpha \leq 2/3$.

The discussion of this subsection is summarized in the following lemma

Lemma 4.3

(i) For $8/7 \leq \alpha \leq 4/3$ the symmetric Nash-Cournot equilibrium price in the neighbourhood of $h_a=h_b$ equals

$$p_i^*(h_i,h_j) = \begin{cases} \dfrac{v+th_i}{2}, & 0 \leq h_i+h_j \leq 2l-\dfrac{2v}{t}, \\[2ex] v-\dfrac{l-h_i-h_j}{2}, & 2l-\dfrac{2v}{t} \leq h_i+h_j \leq \dfrac{l}{2}, \\[2ex] v-th_i, & \dfrac{l}{2} \leq h_i+h_j \leq \dfrac{l}{2}+2h_i-\dfrac{l(v-tl+th_i)}{v-th_i}. \end{cases}$$

(ii) For $8/7 \leq \alpha \leq 4/3$ the symmetric Nash-Cournot equilibrium locations are

$$h_i^* = \frac{v}{t}-\frac{l}{2}.$$

---

[20] Given that $h_i^* \geq l/4$, this condition requires that $\alpha \leq 12/7$. For greater values of α firms have an incentive to increase their price and Economides' touching equilibrium will arise (see also Figure 4.11).

Given the symmetric Nash-Cournot equilibrium locations, the Nash-Cournot equilibrium price equals $tl/2$. For $7/8 \leq \alpha \leq 4/3$ the market equilibrium can then be written as

$$p^* = \frac{\alpha v}{2},$$

$$h^* = \frac{l(2-\alpha)}{2\alpha}.$$

(4.44)

As α drops from $4/3$ to $8/7$, firms will locate *within* the market quartiles. The limit of the Principle of Minimum Differentiation is reached for $\alpha = 8/7$. In that case firm a's location is at $3l/8$, firm b sets up shop at $5l/8$, and the distance between the two firms is only one quarter of the length of the market. At the same time, if the two firms move towards each other (because of lower values of α) they have to lower their price. In particular, the (symmetric) equilibrium price drops from $2v/3$ to $4v/7$ as α goes from $4/3$ to $8/7$.

In the competitive equilibrium with full supply firms always supply their complete hinterland. If then either the length of the market shrinks, the reservation price increases, or the transport cost rises (that is if α falls), the consumer at the market boundary perceives an effective price which is lower than its reservation price. It is then always profitable for a firm to move toward the market centre in order to strive for a larger market share, but never so much that undercutting will be possible.[21] But the closer firms locate (or, the more homogeneous the products are), the fiercer competition will be, and hence, the lower the price they can quote.

Finally, the claim made in the previous subsection, that firms will move away from the market centre towards the market quartiles for $4/3 \leq \alpha \leq 12/7$, needs to be verified. Recall that for this range of α firm i's profits are $\pi_i^*(h_i, h_j) = [v - t h_i][l/2 + h_i - h_j]$. The appropriate partial derivative $\partial \pi_i^*(h_i, h_j)/\partial h_i$ is negative whenever $2h_i - h_j \geq l(2-\alpha)/2\alpha$. Because both firms are located within the market quartiles, we have that $2h_i - h_j \geq l/4$ for symmetric locations. Then observe that $l/4 \geq l(2-\alpha)/2\alpha$ for $\alpha \geq 4/3$. Hence for symmetric locations it is true that $2h_i - h_j \geq l(2-\alpha)/\alpha$ if $\alpha \geq 4/3$, that is, firms move away from each other.

4.6.6 The threat of war ($4/5 \leq \alpha \leq 8/7$)

As explained in the previous sub section, undercutting in the touching equilibrium with full supply will not be profitable as long as $\alpha \geq 4/5$. Firms then have an incentive to move toward each other, which results in the competitive equilibrium with full supply. As we have seen, for symmetric locations undercutting in this equilibrium is not a dominant strategy *iff* $h_i^* \leq 2l/3 - v/3t$. But for $\alpha \leq 8/7$ this

---

[21] Observe that for the range of α under consideration firms do not have to adhere to the strategic ZCV rule for undercutting not to be profitable.

upper limit on either firm's location is smaller than $h_i^* = v/t - l/2$, the Nash-Cournot equilibrium location of the competitive equilibrium with full supply. Hence, for α below 8/7 locations are at $h_i^* = 2l/3 - v/3t$ if we assume that both firms follow the strategic ZCV rule (the equilibrium locations thus follow from the third 'definition' of a local symmetric Nash-Cournot equilibrium).[22] This characterization of the market is then valid as long as $h_i^* = 2l/3 - v/3t \geq l/4$ or $\alpha \geq 4/5$.

The equilibrium price derived in Section 4.6.5 is however not always valid for the values of α under consideration in this sub section. Observe that firms can never form local monopolies if the touching equilibrium with full supply characterizes the market *and* both firms are located at the market boundaries (in terms of Figure 4.10a this means that the situation depicted by prices p_a^0, p_b^0, and locations h_a^0, h_b^0 cannot occur). This is equivalent to $(v - p_a)/t + (v - p_b)/t \geq l$, with $p_i = (v + th_i)/2$, and $h_i = h_j = 0$. This then happens when $\alpha \leq 1$. For these values of α firms are always in competition with each other.

The following lemma summarizes The Threat of War case

Lemma 4.4

(i) For $4/5 \leq \alpha \leq 8/7$ the symmetric Nash-Cournot equilibrium prices in the neighbourhood of $h_a = h_b$ equal

$$p_i^*(h_i,h_j) = \begin{cases} v - \dfrac{l - h_i - h_j}{2}, & 0 \leq h_i + h_j \leq \dfrac{l}{2}, \\[2ex] v - th_i, & \dfrac{l}{2} \leq h_i + h_j \leq \dfrac{l}{2} + 2h_i - \dfrac{l(v - tl + th_i)}{v - th_i}, \end{cases} \quad 4/5 \leq \alpha \leq 1,$$

$$\begin{cases} \dfrac{v + th_i}{2}, & 0 \leq h_i + h_j \leq 2l - \dfrac{2v}{t}, \\[2ex] v - \dfrac{l - h_i - h_j}{2}, & 2l - \dfrac{2v}{t} \leq h_i + h_j \leq \dfrac{l}{2}, \\[2ex] v - th_i, & \dfrac{l}{2} \leq h_i + h_j \leq \dfrac{l}{2} + 2h_i - \dfrac{l(v - tl + th_i)}{v - th_i}. \end{cases} \quad 1 \leq \alpha \leq \dfrac{8}{7},$$

(ii) For $4/5 \leq \alpha \leq 8/7$ the symmetric Nash-Cournot equilibrium locations are

$$h_i^* = \frac{2l}{3} - \frac{v}{3t}.$$

---

[22] Observe that $\partial \pi_i^*(h_i,h_j)/\partial h_i \geq 0$ for all locations $h_i \leq v/t - l/2$, where $\pi_i^*(h_i,h_j)$ are profits in the competitive equilibrium with full supply.

The Nash-Cournot equilibrium price thus equals $4v/3 - 2tl/3$. Consequently, the market equilibrium for $4/5 \le \alpha \le 8/7$ is given by

$$p^* = \frac{2v}{3}(2-\alpha)$$

$$h^* = \frac{l(2\alpha-1)}{3\alpha}.$$

(4.45)

As competition becomes fiercer (that is, when α drops), each firm is forced to move to its hinterland in order not to give the competitor the possibility to undercut. In particular, firm a and b respectively move from $3l/8$ and $5l/8$ to the market quartiles if α falls from $8/7$ to $4/5$. Observe that firms explicitly have to adhere to the strategic ZCV rule for these locations to be equilibrium locations. If not, undercutting becomes a dominant strategy and nothing can be said as to where firms locate in equilibrium. Further, because the respective hinterlands will always be fully supplied, there is room for a price increase if firms relocate away from the market centre. Indeed, if α falls from $8/7$ to $4/5$, the equilibrium price increases from $4v/7$ to $4v/5$.

4.6.7 Making a stand ($0 \le \alpha \le 4/5$)

We finally arrive at the last interval of α to be analyzed. As shown in the previous subsection, for values of α below 1, firms are always in competition with each other. Starting the analysis with both firms located at the market boundaries, we have two candidate equilibria: the touching equilibrium with full supply, and that described by Hotelling. In section 4.6.5 it is shown that touching profits exceed those considered by Hotelling *iff* $3l + h_i - 3h_j \ge 2v/t$. For symmetric locations this translates into $h_i = h_j \le 3l/2 - v/t = l(3\alpha-2)/2\alpha$. Further, for undercutting in the touching equilibrium with full supply not to be profitable, in a symmetric equilibrium it must be that $h_i = h_j \le l/2 - v/5t = l(5\alpha-2)/10\alpha$. Now observe that $(3\alpha-2)/2\alpha \le (5\alpha-2)/10\alpha$ holds for $\alpha \le 4/5$. That is, for $h_i = h_j \le 3l/2 - v/t$, the touching equilibrium with full supply characterizes the market as long as $\alpha \le 4/5$. As shown before, firms then have an incentive to move towards the market centre. But *before* both firms locate at the market quartiles we already have that $h_i = h_j \ge 3l/2 - v/t$, because $\alpha \le 4/5$. Therefore, the competitive equilibrium with full supply will not be reached. Rather, the situation visualized by Hotelling applies. Finally observe that the touching equilibrium with full supply disappears if $3l/2 - v/t \le 0$, or $\alpha \le 2/3$.

Whether or not α is above or below $2/3$, both firms always move towards the market centre. The equilibrium location will then be determined by the undercutting rule, as derived by d'Aspremont et al. [1979]. In particular, if both firms follow their strategic ZCV, in equilibrium they will locate at the market quartiles (see Section 4.2 and 4.6.1).

The analysis can be summarized as follows

Lemma 4.5

(i) For $0 \le \alpha \le 4/5$ the symmetric Nash-Cournot equilibrium prices in the neighbourhood of $h_a = h_b$ equal

$$
p_i^*(h_i, h_j) = \begin{cases}
v - \dfrac{l - h_i - h_j}{2}, & 0 \le \dfrac{4l(h_i + h_j)}{3} \le \dfrac{4l(3l - 2v/l + 4h_i)}{9}, \\[3mm]
t\left(l + \dfrac{h_i - h_j}{3}\right), & \dfrac{4l(3l - 2v/l + 4h_i)}{9} \le \dfrac{4l(h_i + h_j)}{3} \le \left(l + \dfrac{h_i - h_j}{3}\right)^2, & \dfrac{2}{3} \le \alpha \le \dfrac{4}{5} \\[4mm]
t\left(l + \dfrac{h_i - h_j}{3}\right), & 0 \le \dfrac{4l(h_i + h_j)}{3} \le \left(l + \dfrac{h_i - h_j}{3}\right)^2, & 0 \le \alpha \le \dfrac{2}{3}.
\end{cases}
$$

(ii) For $0 \le \alpha \le 4/5$ the symmetric Nash-Cournot equilibrium locations are

$$
h_i^* = \frac{l}{4}.
$$

4.6.8 Mixed strategy equilibria

Before we can conclude that we have characterized the equilibrium (for symmetric locations) of Hotelling's spatial model with an outside good, we should consider firms' incentives (if any) to enter the undercutting area if they employ mixed pricing and/or location strategies. To the extent that firms adhere to strategic ZCV these strategies become redundant, for in this case neither firm will give its rival the opportunity to undercut. Still, mixed strategies in Hotelling's original framework have been thoroughly studied, the results of which deserving to be discussed here.

At forehand however it is difficult to imagine that plant locations are chosen randomly. As Martin [1993, p.274] puts it:

> Real-world decisions about the locations of plants are not made randomly. Models that possess pure-strategy equilibria are more likely to yield *a priori* plausible hypotheses than models with mixed strategy equilibria.

This view is shared by de Palma et al. [1985, p.768]:

> ...intuitively, mixed strategies do not have sufficient predictive power to account for most locational decisions.

On the other hand, many prices do follow erratic patterns, which could well be described by mixed pricing strategies. We therefore restrict ourselves to situations in which strategies in locations are pure, while those in prices are mixed.

A result by Dasgupta and Maskin [1986] ensures that a Nash-Cournot equilibrium price exists for any pair of locations if mixed strategies in the second stage of the game are allowed. But using a mixed strategy in price might affect the (pure strategy) location decision, as Friedman [1988] points out (p.616, word in parentheses added):

> I see no way to rule the possibility that the equilibrium (..) will be destroyed if mixed strategies are allowed. Perhaps a firm might increases its expected payoff by switching to an output level (location) whose second-stage best play is a mixed-price choice. Analyzing this question appears extremely formidable.

Encouraged by the result of Dasgupta and Maskin [1986], Osborne and Pitchik [1987] try to characterize the equilibrium of Hotelling's model for mixed pricing strategies, which indeed appears to be 'extremely formidable'. First they prove that an equilibrium with mixed pricing strategies does not exist if (i) $(1+(h_i-h_j)/3)^2 \geq 4(h_i+2x_j)/3$, or (ii) $h_a+h_b=1$. These are exactly the locations for which d'Aspremont et al. [1979] establish the pure strategy Nash-Cournot price equilibrium. Second, using simulations, Osborne and Pitchik [1987] find a unique and symmetric Nash-Cournot location equilibrium such that $0.266l < h_i^* < 0.274l$. Incidently, these locations are almost at the market quartiles, the equilibrium locations corresponding to the Principle of Intermediate Differentiation.

Whether or not firms would indeed locate within the market quartiles depends on expected profits earned at these locations. The Nash-Cournot equilibrium price derived by Osborne and Pitchik [1987], defined over some probability density function, never exceeds the Hotelling price for symmetric locations (tl). Hence, profits when allowing for mixed strategies in price never exceed those obtained when both firms use pure pricing strategies and follow their strategic ZCV. That is, although firms can move (a bit) closer towards each other when they randomly set price, they will not do so in order not to lose profits. As a result, the equilibrium derived above for $\alpha \leq 2/3$ is indeed the two stage Nash-Cournot equilibrium, provided that firms adhere to strategic ZCVs.

It is tempting to conclude the same for values of α exceeding $2/3$, given the enormous effort of Osborne and Pitchik [1987] in a relatively simple setting. For $\alpha \geq 8/7$ this seems quite harmless, because the equilibrium locations for this range of α are determined by profit maximization considerations. Allowing for mixed prizing strategies might establish an equilibrium in location for those locations where undercutting is profitable in the pure strategy game (that is, when the distance between the two firms is small). But it is not obvious that, while firms want to move away from each

other when using pure pricing strategies, they remain at the same place (or locate even closer to each other) when using mixed pricing strategies. And even if so, it is unlikely that the expected profits are then greater than those obtained in the pure strategy equilibrium, given that in the latter case it is profitable to move away from each other. Further, if $2/3 \leq \alpha \leq 4/5$ we can rely on Osborne and Pitchik [1987]. For this range of α the market segment where undercutting becomes profitable is exactly the same as that in Hotelling's model, and the reasoning of the previous paragraph applies. But for $4/5 \leq \alpha \leq 8/7$ an analysis similar to that of Osborne and Pitchik [1987] could be repeated. Indeed, here we explicitly rely on both merchants having strategic ZCV, thereby skirting the possibility of undercutting. To take this into account would require an effort beyond the scope of this study.

4.7 The principle of almost intermediate differentiation

The analysis of Section 4.6 gives rise to a symmetric equilibrium price and concomitant equilibrium location for all values of α. These are summarized in Proposition 4.1.

The equilibrium price is illustrated in Figure 4.14 as the locus ODEFC. In the same figure also two other benchmarks are depicted: the locus OABC is the equilibrium price if both firms are constrained to be at the endpoints of the linear market (this is equivalent to Beckmann and Thisse [1986, p.45, Figure 3.5]), while the locus ODC depicts the equilibrium price if the locations of the two firms are constrained to be at the market quartiles.

Observe that if firms compete the first mentioned constrained equilibrium price (locus OABC) is below that when firms are able to move freely. If the reservation price falls, either firm can gain profits by moving inwards. But if this is not possible, then the gain in profits due to a larger market share exceeds the loss in profits because of a lower price. Hence, a net increase in profits can be expected from a lower price. In fact, for values of α between 0 and 1, the constrained equilibrium price is such that both firms together just cover the entire market. For larger values of α some customers in the middle of the market will not be served.

On the other hand, if both merchants are constrained to be located at the market quartiles, the only difference from the unconstrained equilibrium is for that range of α where it is profitable for either firm to move inwards (that is for $4/5 < \alpha < 4/3$). But because now firms cannot relocate, they are able to charge a higher price due to weakened price competition.

The equilibrium location is depicted in Figure 4.15 as locus ABCDE plus the shaded monopoly locations (compare Figure 4.4). The locus ABCFGH divides the linear market in second stage pure strategy equilibria (to the left) and mixed strategy equilibria (to the right). The arrows indicate the relocation tendencies of producers following the sign of the derivative of first stage equilibrium profits with respect to location.

Leaving aside the local monopoly case, it is immediate that firms will never move closer together than a quarter of the length of the market, and never further apart than half the length of the market. For most values of α however both firms locate at the market quartiles if they actually compete. Hence it would be appropriate to refer to a Principle of Almost Intermediate Differentiation:

firms have a tendency to move closer together to serve a larger market, but tend to retreat because of fiercer competition.

Proposition 4.1

Given that $l=\alpha v/t$, the symmetric Nash-Cournot equilibrium price, p^*, and location, h^*, are respectively given by

$$
p^* = \begin{cases}
tl, & 0 \leq \alpha \leq \dfrac{4}{5}, \\[2ex]
v\left(\dfrac{2(2-\alpha)}{3}\right), & \dfrac{4}{5} \leq \alpha \leq \dfrac{8}{7}, \\[2ex]
v\dfrac{\alpha}{2}, & \dfrac{8}{7} \leq \alpha \leq \dfrac{4}{3}, \\[2ex]
v\left(\dfrac{4-4\alpha}{4}\right), & \dfrac{4}{3} \leq \alpha \leq 2, \\[2ex]
\dfrac{v}{2}, & 2 \leq \alpha < \infty,
\end{cases}
$$

and

$$
h^* = \begin{cases}
\dfrac{l}{4}, & 0 \leq \alpha \leq \dfrac{4}{5}, \\[2ex]
\dfrac{l(2\alpha-1)}{3\alpha}, & \dfrac{4}{5} \leq \alpha \leq \dfrac{8}{7}, \\[2ex]
\dfrac{l(2-\alpha)}{2\alpha}, & \dfrac{8}{7} \leq \alpha \leq \dfrac{4}{3}, \\[2ex]
\dfrac{l}{4}, & \dfrac{4}{3} \leq \alpha \leq 2,
\end{cases}
$$

$$
h^* \in [\dfrac{1}{2\alpha}, \dfrac{\alpha-1}{2\alpha}], \quad 2 \leq \alpha < \infty.
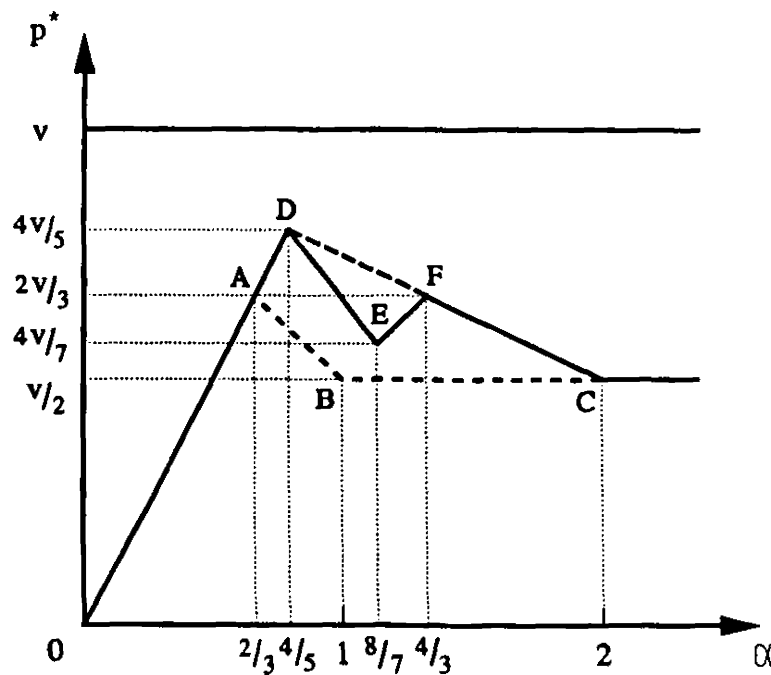$$

Figure 4.14 equilibrium prices

4.8 Conclusions

More than half a century ago Harold Hotelling introduced space into micro-economic theory. His model of spatial duopoly is seminal and for many years economists believed the Principle of Minimum Differentiation he derived from it. D'Aspremont et al. [1979] revealed a fundamental flaw in Hotelling's analysis and showed four years later that the Principle of Minimum Differentiation never holds in a Hotelling-like world. Since then many economists have tried to achieve stability in competition by relaxing some of Hotelling's (implicit) assumptions. This chapter discussed one such generalization: endowing each consumer with a finite reservation price.

If consumers have an upper bound on the price they are willing to pay for a horizontally differentiated product, quite different market equilibria emerge from a simple spatial duopoly than that predicted by the Principle of Minimum Differentiation. If the market is large relative to the effective reservation price, both firms can form local monopolies. Locations are than undetermined, up to a certain interval. On the other hand, if the size of the market, relative to the effective reservation price, is such that firms compete, they will never move closer together than a quarter of the length of the market, and never further apart than half the length of the market, leading to the Principle of Almost Intermediate Differentiation.

We started this chapter by observing that economic theory is still far from providing a unanimous answer as to where two competing firms will locate their production facilities. Depending on what assumptions are (not) being made, either it is predicted that firms will locate back-to-back,
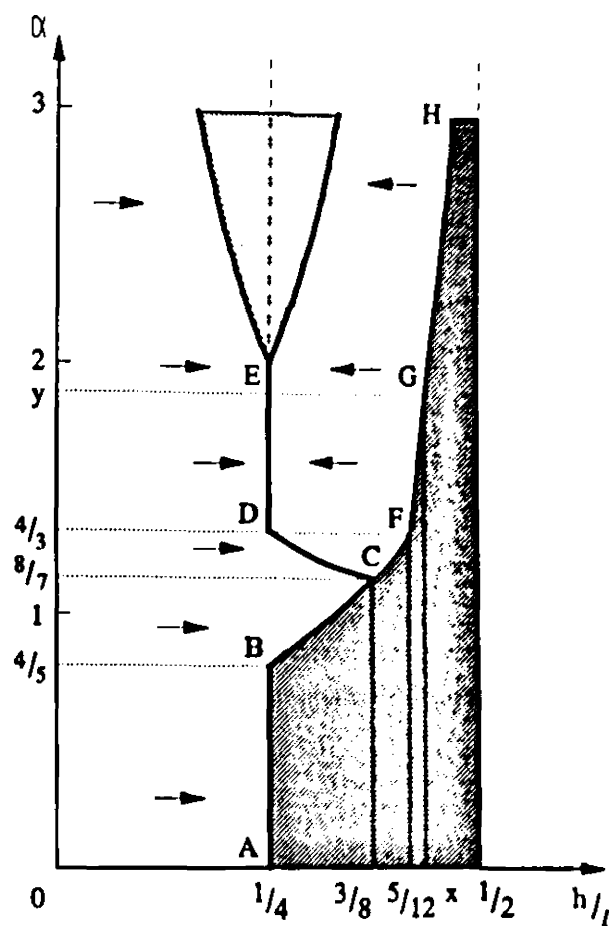
Figure 4.15 Equilibrium locations; x ~ 0.43, y = 6(4+√10)/(7+5√10)

as far as possible from each other, or somewhere in between. The analysis presented in this chapter revealed that assuming each consumer to have an infinite reservation price (an assumption that is being made in the bulk of contributions to the area of spatial economic theory) is not a harmless assumption. It affects the price-location equilibrium dramatically, and many analyses could be reconsidered in the light of finite reservation prices. Indeed, the location puzzle is still far from being solved.

Appendix 4A Undercutting in Economides' competitive equilibrium: small markets

First we have to determine *when* markets are small. Recall that in Economides' touching equilibrium undercutting is not profitable as long as

$$l - h_i - h_j > \frac{6v}{t(7 + 5\sqrt{10})}.$$ (4A.1)

This condition was derived under the 'large market' assumption

$$h_i \geq \frac{v - p_i^u}{t}.$$ (4A.2)

If this assumption is not met, profits when undercutting are, of course, lower compared to the situation considered in Section 4.5.2. Therefore, for undercutting to be profitable in 'small markets' it is necessary that

$$l - h_i - h_j \leq \frac{6v}{t(7 + 5\sqrt{10})}.$$ (4A.3)

Then, given this condition and the fact that $p_i^u = 2[v - 2t(l - h_i - h_j)]/5$, (4A.2) translates, for symmetric locations, into

$$\frac{l}{2} - \frac{3v}{t(7 + 5\sqrt{10})} \leq h_i = h_j \leq \frac{3v}{5t} + \frac{4}{5t}\left(\frac{6v}{7 + 5\sqrt{10}}\right),$$ (4A.4)

or

$$\alpha \leq \frac{6(4 + \sqrt{10})}{7 + 5\sqrt{10}}.$$ (4A.5)

Second, given that markets are small, we have to establish when it is profitable to undercut. Remember that in absence of undercutting, Economides' competitive profits are

$$\pi_i^c(h_i,h_j) = \frac{3}{2t}\left(\frac{2v+t(l-h_i-h_j)}{5}\right)^2.$$

(4A.6)

On the other hand, profits when undercutting equal

$$\pi_i^u(h_i,h_j) = p_i^u(h_i,h_j)\left(h_i+\frac{v-p_i^u(h_i,h_j)}{t}\right).$$

(4A.7)

Using some tedious algebra it can be derived that

$$\pi_i^c(h_i,h_j) \geq \pi_i^u(h_i,h_j) \quad \Leftrightarrow \quad l-h_i-h_j \leq \frac{2}{7}\left(\sqrt{\left(\frac{v}{t}+2h_i\right)^2+7\frac{v}{t}h_i} \ -\frac{v}{t}-2h_i\right),$$

(4A.8)

the condition given in the text.

Part III

Robust Estimation

It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

- Sherlock Holmes -

5   General M estimation

5.1 Introduction

Empirical observations not in line with prior expectations or concomitant empirical evidence have attracted special attention as long as data have been recorded. In 1777 the Dutch astronomer Daniel Bernoulli writes (Allen [1961])

> is it right to hold that the several observations are of the same weight or moment, or equally prone to any and every error?...Is there everywhere the same probability? Such an assertion would be quite absurd, which is undoubtedly the reason why astronomers prefer to reject completely observations which they judge to be too wide of the truth, while retaining the rest and, indeed, assigning to them the same reliability.

Indeed, thorough examination of data before statistical techniques are applied, is a tradition in empirical research which goes back centuries. When Adrien Marie Legendre published in 1805 his *méthode des moindres quarrés* for example, he also discusses the Ordinary Least Squares (OLS) residuals (Plackett [1972])[1]

> If among these errors are some which appear too large to be admissible, then those observations which produced these errors will be rejected, as coming from too faulty experiments, and the unknowns will be determined by means of the other observations, which will then give much smaller errors.

Both Bernoulli and Legendre refer to the use of some *robust* statistical procedure to draw conclusions. First, the data have to be carefully examined, and observations which appear to be outlying, according to some rule, have to be rejected. Second, apply statistical techniques on these 'cleaned' data.

Although the occupation with anomalies in data is as old as statistical research, it was not until 1964 that a formal treatment of robust estimation saw the light. In his seminal paper "Robust estimation of a location parameter", Huber [1964] introduced a robust estimator (and derived its asymptotic properties) for estimating a location parameter based in a sample of observations drawn from a contaminated normal distribution. With this paper, Huber paved the way for the development of robust estimation theory as it is known today.

It is surprising to observe however that more than 30 years after the seed for robust estimation was planted, and a sophisticated apparatus for estimating unknown parameters robustly has been developed, many applied researchers still do not use these techniques. Indeed, bridging to some extent the gap between formal statistics and applied econometrics, by deriving and applying a particular robust estimator for the linear model, is one of the aims of this and the next chapter.

---

[1] Legendre and Carl Friedrich Gauss fought a life long combat for the recognition of having invented OLS. For an entertaining overview of this battle, including some recent evidence supporting Gauss' claim, see Stigler [1981].

As we shall see, there are many reasons why outlying observations (that is, observations which appear to be inconsistent with the remainder of the set of data (Barnett and Lewis [1984, p.4])) are present in any data set. In fact, it is almost impossible to construct an empirical dataset without any outliers. Hence, the statistician confronted with 'real' data always has to make the following decision: either the outlying observations are left untreated and traditional statistical analyses are carried out on the raw data (with often dramatic consequences!), or the outliers are given special consideration. In the latter case, one of two approaches can be followed. The first is to *accommodate* the outliers. This means that all data are used in the statistical analysis, but the influence of outlying observations is controlled for. Robust estimation techniques are examples of this accommodation approach.

The second way to handle data anomalies is the two-step procedure advocated by the illustrious scientists as cited above. This second approach has two main setbacks however. First, in many cases, outliers are defined relative to the unknown values of the model parameters. Hence, these have to be estimated first (robustly) in order to identify the outlying observations. Second, accommodating outliers is a smooth transition between full acceptance and full rejection of the outlying observations. As a result, robust procedures are (much) more efficient than rejection procedures. Indeed, outliers can well be genuine and important observations. Down-weighting these points appropriately retains to some extent the information they carry, without allowing them to influence too much the overall estimate.

In this chapter a robust estimator, a so-called GM estimator, is developed for the unknown parameters of the linear model. In applied statistical research it is common practise to use OLS-like techniques to estimate these unknowns. However, as will be illustrated below (and as should be common knowledge!), these classic estimation techniques are highly sensitive to outlying observations. For instance, any single observation can cause OLS to yield any estimate with any level of significance (for some prominent examples of this vulnerability see Hinloopen and Wagenvoort [1995] and Section 5.9.3 below).[2] This weakness of classic estimation techniques features all the more prominently in an era where proper data examination is often the exception rather than the rule (or, at least, is not reported). Yet another purpose of the present and subsequent chapter is to examine to what extent this practise is responsible for statistical digressions. In particular, in Chapter 6 the GM estimator is applied to existing data sets, and the results are compared with those published previously.

The plan of action for the current chapter is as follows. First, outlying observations are examined more carefully. That is, it is examined how they emerge, what they look like, and how critical they are. Some important concepts, the breakdown point and the influence function, are introduced in Section 5.3, followed by an outline of the estimator we have developed. Sections 5.5 through 5.8 deal with technical details concerning the GM estimator, while in section 5.9 its statistical properties, such as consistency and efficiency, are discussed. Section 5.10 concludes.

---

[2] An early attempt to develop a robust estimator, in the spirit of OLS, was made by Edgeworth in 1887. Instead of minimizing the sum of *squared* errors, Edgeworth [1887] proposed to minimize the sum of *absolute* errors, giving rise to the Least Absolute Values estimator. Although this estimator is less sensitive than OLS to certain types of outliers (see e.g. Judge et al. [1988, p.899]), it can still be tricked by a single outlying observation (see e.g. Rousseeuw and Leroy [1987, p.10-11]).

5.2 Outliers: birth, species and importance

Throughout this chapter we are concerned with estimating the unknown parameter vector β in the linear model

$$y_i = x_i\beta + e_i, \qquad i = 1,...,n, \tag{5.1}$$

where $y = \{y_1,...,y_n\}^T$ is an observable dependent variable, x_i a row vector of length p of observable explanatory variables, and the errors, e_i, are independent and identically distributed with zero mean and bounded variance (unless stated otherwise, the fact that there are n observations is implicit in this chapter). To be more specific, we are concerned with the effects of outlying observations in both y and $X = \{x_1^T,...,x_n^T\}^T$ on the estimate b of β in (5.1). For that, first the origin, the different types, and the importance of outliers are to be examined in more detail.

5.2.1 The birth of an outlier

During the course of empirical research, one of several things can go wrong, causing the experimenter to be confronted with outlying observations. In this subsection the four main sources of data contamination will be discussed.[3]

The most blatant origin of data corruption is human error. Typos, copying errors, wrong computations, and transcription errors all could result in obscure observations (indeed, many statisticians consider these observations to be the only true outliers). The reader might conjecture that careful scientific practise precludes this type of mistake. Alas, in Section 6.3 a conspicuous transcription error is revealed in a study published just recently in an influential economics journal.

Related to these human errors is ignorance. Examples abound for instance that observers did not realize a change in scale of measurement. Barnett and Lewis [1984, p.7] illustrate what happens if a change in temperature scale from degrees Fahrenheit to degrees Celsius remains unnoticed. And if an interchange between the time length of a unit of observation and the number of observations remains undetected, quite dramatic outliers emerge, as illustrated by Rousseeuw and Leroy [1987, p.25-26]. Of course, these examples exist since the ignorance, leading to the creation of outliers, was revealed through careful examination of the appropriate data. Often, however, this important step is left out of the statistical analysis.

Bad or outdated equipment and unskilled observers can both lead to measurement errors, the third seed for outlying observations. On a more fundamental level, especially in social sciences, the unit of observation is often only a proxy for the unit of interest. Hence, even if the observation is measured without error, a fundamental measurement error is still unavoidable. Outlying observations then arise if the observed variable is more volatile than the (theoretical) variable of interest.

This brings us to the last main reason as to the birth of outlying observations. It could be that the model imposed on the data *a priori* is different from that underlying the data. For instance, if a normal distribution is imposed on observations drawn from a Cauchy distribution, several observations will be considered to be located too far from the sample mean (or median), thus unjustly being

---

[3] This subsection is based on Barnett and Lewis [1984] and Rousseeuw and Leroy [1987].

identified as outliers. In applied economic research something similar could happen if the model to be estimated is taken from economic theory. A fundamental problem then arises: should the model be augmented to fit the data (if necessary), or should the data be augmented to fit the model (if necessary)? Of course, no univocal answer is to be given to this dilemma. If the data are poor and contain clear 'bad' observations, the original model should prevail. On the other hand, if many empirical studies reject certain models, it is difficult to retain these. As it stands, in applied economic research it is common practise to find empirical justification for models rooted in economic theory. The possible construction of outlying observation thus features prominently.

Having examined many empirical studies in 'hard' sciences, Hampel et al. [1986] conclude (Hampel et al. [1986, p.28], word in parentheses added)

> We have to accept the whole range of data quality, as well as the different causes of outliers...altogether, 1-10% gross errors (outliers) in routine data seem to be more the rule rather than the exception.

Also Barnett and Lewis [1984] implicitly claim that outlying observations are present in any empirical data set.

5.2.2 The family of outliers

If outlying observations are to be examined, it is useful to categorize them first, since different species may deserve different treatment. Within the world of the linear model, two types of outliers are to be distinguished. In defining these we roughly follow Rousseeuw and Van Zomeren [1990].

Definition 5.1 Leverage Points

An observation is a leverage point if, within the hyper plane of explanatory variables, it lies far from the bulk of explanatory variables.

Observe that leverage points only concern the regressors x_i. That is, the identification of a leverage point is independent of the fitted linear relation between X and y. Note in passing that for the identification of leverage points according to Definition 5.1, both 'far' and 'the bulk' have to be made explicit. We will come back to this in extensive detail in Sections 5.6 and 5.7. Meanwhile realize that leverage points as such can hinder their identification if traditional measures are used. For instance, both the arithmetic mean and concomitant variance are influenced heavily by leverage points, thus obscuring the classic z-score.[4]

Data which deviate from the linear pattern fitting the majority of observations are labelled 'regression outliers' by Rousseeuw [1984]. This notion of outlyingness for the response variable does not take into account however whether or not the observation is also a leverage point. To distinguish between regression outliers which are also leverage points, and regression outliers which are not, Rousseeuw and Van Zomeren [1990] introduce the notion of a vertical outlier.

---

[4] See equation (5.24) and the accompanying text for a discussion of the z-score.
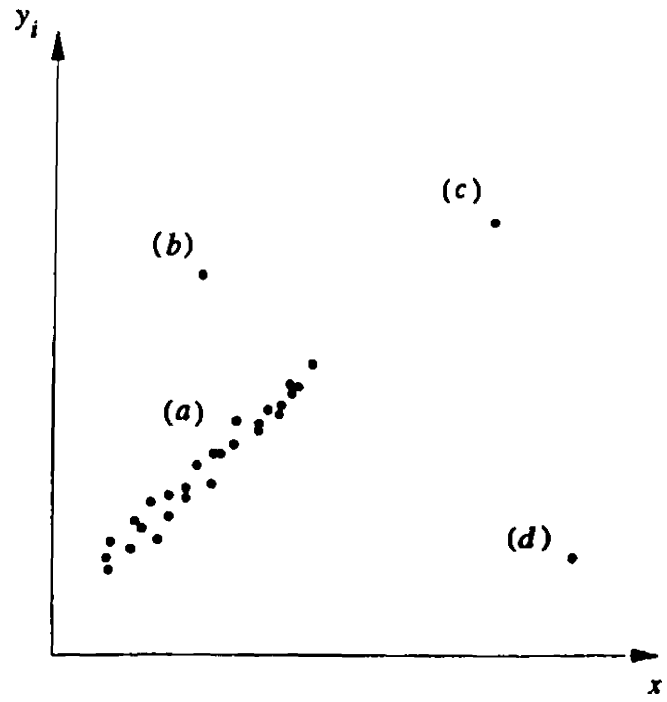
Figure 5.1   (*a*) regular observations, (*b*) vertical outlier, (*c*) leverage point, and (*d*) leverage point and vertical outlier

Definition 5.2 Vertical Outliers

An observation is a vertical outlier if it deviates from the linear pattern of the majority of the data but is not necessarily a leverage point.

Observe that vertical outliers are defined relative to the fitted linear relation between X and y. The salient features this brings about for a robust estimation procedure are discussed in Sections 5.4 and 5.8.

In Figure 5.1 the different types of outlying observations are illustrated for a simple linear regression, relative to the regular data (*a*). Observations (*b*) and (*d*) are both vertical outliers and thus both regression outliers. However, (*d*) is also a leverage point whereas (*b*) is not. Observation (*c*) on the other hand is only a leverage point.

5.2.3 The importance of outliers

Outlying observations are by no means equivalent to 'bad' observations. Depending on the type of problem a researcher is occupied with, outliers should be down-weighted (possibly to zero) or fully taken into account. Indeed, situations abound in which the most valuable information is contained in those observations not following the majority of the data. Engineers, for example, involved with

Figure 5.2   Daily water level river Maas at Borgharen dorp, 1990 - 1993

determining the height of a dyke have little interest in the average water level. Rather, the extreme water levels are of fundamental interest, since it is these heights the dyke should encompass.
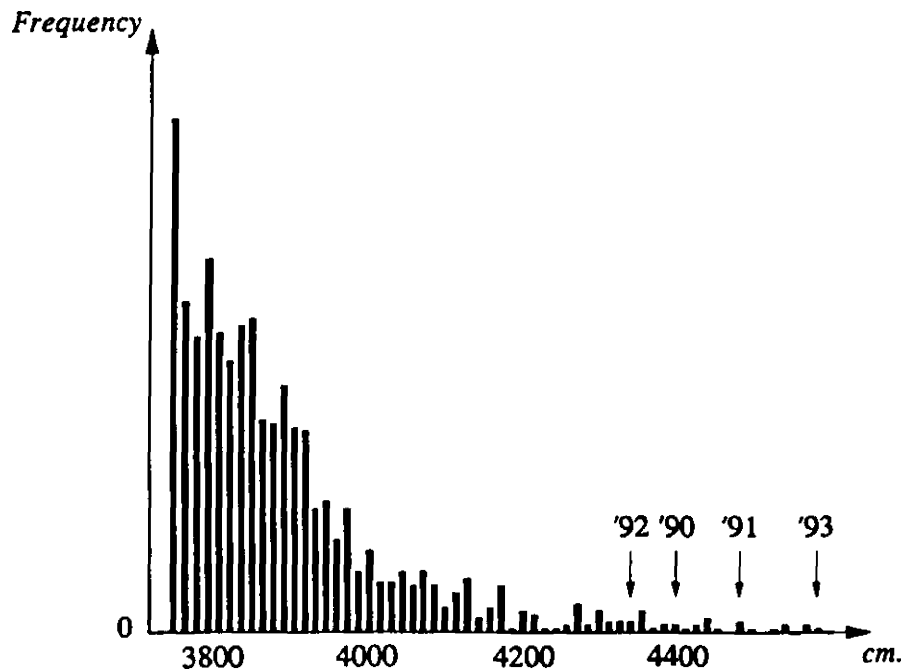
To illustrate the importance of outlying observations, in Figure 5.2 the daily levels of the river Maas at Borgharen dorp (The Netherlands) is depicted for four years, starting in 1990.[5] The annual maximum water levels are also indicated (these were 4,000, 4,482, 4,336 and 4,585 centimetres respectively). If for example the hight of the dyke is set at the median water level plus 5 times the standard deviation (a height that could be considered to encompass all conceivable heights), and only the observations until 1993 are known, it would reach a mere 4,560 centimetres. But then, in 1993, on two occasions (outliers!) the water would have risen above the dyke.[6]

That the importance of an outlier is often only a *relative* concept however, can also be illustrated with the water level data. Again, if only the observations until 1993 are considered, a maximum water level of 4,482 centimetres would be registered (the maximum of 1991). This observation has the highest z-score (4.50), relative to the restricted sample, and could be considered very important in determining the height of the dyke. However, if also the data concerning 1993 are included, the maximum water level of 1991 becomes a less important observation. Not only does its z-score drops to 4.16, making it a less extreme outlier, but there are now 7 observations with a higher z-score, 5 of which are above 4.50.

---

[5] These data were kindly provided by ir. Chbab of the 'Ministerie van Verkeer en Waterstaat'.

[6] In fact, this is exactly what happened. Yet it took another flooding, in 1994, for policymakers to understand that the dykes should be heightened to guard the population against future outliers.

In economics on the other hand, researchers are often concerned with the statistical pattern exhibited by the majority of the data. Outlying observations should of course be identified, since they could reflect important occurrences. But if structural relations in experimental data are to be revealed, outlying observations could be misleading. Indeed, it is precisely the relation exhibited by the majority of the data which are revealed by robust estimation techniques, while at the same time identifying those observations which deviate from this pattern.

5.3 Some statistical concepts

Several indicators have been developed to assess the (relative) performance of an estimator. Among these are the extent to which, on average, the estimate differs from the true value of the unknown parameter (bias), its relative variance (efficiency) and the idea that it converges (in probability) to the true value of the unknown parameter if the sample size increases (consistency). Recently, especially in the context of data contamination, additional statistical concepts have been proposed to value the quality of an estimator. In this section two of the most widely used among these are discussed: the influence function and the breakdown point.

5.3.1 Influence function

The influence curve or influence function measures the impact of a single observation on a statistic. The curve is due to Hampel [1968] and (Hampel [1974, p.383])

is essentially the first derivative of an estimator, viewed as a functional, at some distribution.

More formally, let $z = (z_1, ..., z_n)^T$ be a sample of observations drawn from some distribution F. To assess the influence of an observation on a statistic, $T(\cdot)$, defined over z, we can simply add one observation to the sample. Let the additional observation, z^*, be drawn from an atomistic distribution G, with $P_r(g = z^*) = 1$. The influence function of $T(z)$ under F is then defined as

$$IF_{T,F}(z^*) = \lim_{\delta \downarrow 0} \{ [T([1-\delta]F + \delta G) - T(F)]/\delta \}, \tag{5.2}$$

or, equivalently

$$IF_{T,F}(z^*) = \frac{\partial}{\partial \delta} T([1-\delta]F + \delta G)|_{\delta=0}. \tag{5.3}$$

The appeal of the influence function as a characterization of a statistic is its heuristic interpretation. It describes for each point z_i the effect of an infinitesimal contamination in z_i on the value of the statistic. Clearly, a desirable feature of any statistic is that this effect is bounded, given that any dataset contains at least some outliers (see Section 5.2.1). That is, the influence function should be bounded. In order to examine this alleged property, the supremum of the absolute value of (5.2) or (5.3) over z could be determined; *the gross-error*. Observe that this supremum measures the worst influence an infinitesimal contamination could have on the value of the concomitant statistic.

Example 5.1 The influence function of OLS

Let $x = \{x_1,...,x_n\}^T$ and $y = \{y_1,...,y_n\}^T$ be drawn from a distribution F and suppose that the observation $\{x^*,y^*\}$, drawn from the atomistic distribution G, is added to this sample. The OLS-rule for estimating β in (5.1) applied to $(1-\delta)F+\delta G$ yields

$$b^{OLS}([1-\delta]F+\delta G) = \frac{(1-\delta)\sum_{i=1}^{n} x_i y_i + \delta x^* y^*}{(1-\delta)\sum_{i=1}^{n} x_i^2 + \delta x_i^{*2}}. \tag{5.4}$$

Inserting (5.4) into (5.3), taking the appropriate derivative and simplify finally results

$$IF_{T,F}^{OLS}(x^*,y^*) = x^*[y^* - x^* b^{OLS}(x,y)] \Big/ \sum_{i=1}^{n} x_i^2. \tag{5.5}$$

From (5.5) we can make two observations. First, the gross-error of OLS goes to infinity, since the difference between y^* and $x^* b^{OLS}(x,y)$ can become arbitrarily large. In other words, the influence of a small amount of contamination could be unbounded. Second, the influence of the contamination, $\{x^*,y^*\}$ consists of two components: the relative size of the independent variable ($x^*/\sum_i x_i^2$) and the associated residual ($y^* - x^* b^{OLS}(x,y)$). It is precisely these two channels of influence which the estimator presented in the next section tries to dampen.

5.3.2 Breakdown point

Another property of an estimator is the percentage of data contamination that leaves it in some sense undisturbed. To examine this feature more precisely Hampel [1971] has introduced the breakdown point: the smallest fraction of data contamination that can cause an estimator to take on arbitrary values.[7] Formally, let $\eta(\gamma;T(\cdot),X)$ be the supremum of $|T(X') - T(X)|$ for all X', where X' corresponds to the original data set with a fraction γ ($=l/n$) of the observations replaced by arbitrary values, and where $|\cdot|$ is some matrix norm. The breakdown point of a statistic $T(\ )$ on X is then defined as

$$\varepsilon^*[T(\cdot),X] = \min\left\{\frac{l}{n};\eta(l;T(\cdot),X) = \infty\right\}, \tag{5.6}$$

where l is the number of original data points replaced by arbitrary values. Obviously, a statistic's breakdown point should be as high as possible. However, by definition 50% is the highest breakdown point achievable, since beyond this limit the distinction between 'good' and 'bad' observations becomes arbitrary.

---

[7] The definition we use here follows the finite sample version of Donoho and Huber [1983].

Example 5.2 The breakdown point of OLS

Consider the linear model (5.1). In order to examine the breakdown point of OLS, l observations are replaced by arbitrary values. Let I_l be the set of indices corresponding to these observations and denote perturbed values by a star. In particular, $x_i^* = (1 + \lambda_i)x_i$ and $y_i^* = (1 + \mu_i)y_i$, $\lambda_i, \mu_i \in \mathbf{R}$, $i \in I_l$. Define $X = \{x, y\}$ and $X' = \{x_i, x_j^*, y_i, y_j^*\}$, $i \in I_{n-l}$, $j \in I_l$. Applying the OLS-rule for estimating β in (5.1), given the contaminated data, set leads to

$$b^{OLS}(X') = \frac{\sum_i x_i y_i + \sum_{i \in I_m} x_i \mu_i + y_i \lambda_i + \lambda_i \mu_i}{\sum_i x_i^2 + \sum_{i \in I_m} \lambda_i(x_i + \lambda_i)}. \tag{5.7}$$

Clearly, any *single* λ_i or μ_i can lead to a difference between $b^{OLS}(X)$ and $b^{OLS}(X')$ that goes to infinity. In other words, l equals 1 in (5.6), implying that the breakdown point of OLS is $1/n$. This fraction goes to 0 for arbitrarily large n.

Having described important additional means to assess the performance of a statistic, and having illustrated that OLS performs particularly badly according to these statistical indicators, we turn to the construction of an estimator which combines reliable performance according to the robustness measures with good traditional statistical properties.

5.4 Multi-step GM estimation

As illustrated in Example 5.1, the influence of an infinitesimal contamination on the OLS estimate of β in (5.1) is felt in two ways: via the residual and through the relative size of the contaminated explanatory variable. In this section an estimator is derived which controls these two channels of influence.

A first step in this direction is to diminish the impact of vertical outliers. This is achieved by appropriately reformulating the objective function associated with estimating β in (5.1). In particular, Huber [1964] proposed in his seminal paper the objective

$$\min_b \sum_{i=1}^{n} \rho(r_i), \tag{5.8}$$

where $r_i = y_i - x_i b$, and $\rho(\cdot) : \mathbf{R} \to \mathbf{R}_+$, with $\forall t \in \mathbf{R}$, $\rho(t) = \rho(-t)$, and $\rho(0) = 0$ (in case of OLS we have that $\rho(r_i) = r_i^2$).[8] The corresponding first order condition reads as

---

[8] In the literature, some additional regularity conditions are imposed on $\rho(\cdot)$ (see Hampel et al. [1986, p.315]).

$$\sum_{i=1}^{n} x_i^T r_i w_r(r_i/\sigma_r) = 0, \tag{5.9}$$

where x_i^T is the transpose of x_i, where the weight function $w_r(\cdot): \mathbf{R} \to [0,1]$ is by definition given by $w_r(r_i/\sigma) = [\partial\rho(r_i/\sigma)/\partial r_i]/r_i = \psi(r_i/\sigma)/r_i$, and where σ_r is the variance of the residuals (the residuals are standardized, since otherwise the solution to (5.9) would not be equivariant with respect to a rescaling of the y-axis; see also Section 5.5.2). Note that $w_r(\cdot)$ down-weights vertical outliers.

The estimator implicitly defined by (5.9) belongs to the class of *M estimators*. The Ordinary Maximum Likelihood estimator, which is obtained by defining $\rho(r_i)$ as $-\ln[\phi(r_i)]$, where $\phi(\cdot)$ is the probability density function of r_i, is a member of this family.

Although the influence of a residual on the estimate of β in (5.1) is bounded if $\rho(\cdot)$ is bounded, it is clear from (5.9) that the second channel through which the effect of data contaminations flows, that is, the relative size of the contaminated explanatory variable, is still left unattended. Indeed, as shown by Hampel [1973], the influence function of an M estimator can be written as the product of two factors, being the *influence of the residual*, and the *influence of position in the design matrix*. Since the latter is not controlled for, the solution to (5.9) remains vulnerable to leverage points. As a result, the breakdown point of M estimators is still only $1/n$ (see e.g. Rousseeuw and Leroy [1987]).

Accordingly, Mallows [1975] proposed to replace (5.9) by

$$\sum_{i=1}^{n} x_i^T w_x(x_i) r_i w_r(r_i/\sigma_r) = 0, \tag{5.10}$$

where $w_x(\cdot): \mathbf{R}^p \to [0,1]$ is a weight function based on the design matrix X. The estimator implicitly defined by (5.10) is a so-called *Generalized M* or bounded influence estimator, since both leverage points and vertical outliers are down-weighted.

Alas, as shown by Maronna et al. [1979], the breakdown point of a GM estimator for β in (5.1) is at most $1/(p+1)$. For any p of reasonable size this is well below 50%.

It took more than a decade before the problem of the low breakdown point of GM estimators was fixed. Observe that, since (5.10) does not have a closed form solution, numerical procedures have to be used to solve the first order condition defining the GM estimate. Indeed, the solution to the low breakdown point problem lies precisely in this numerical procedure. In a seminal paper Simpson et al. [1992] carefully point out how to construct GM estimators which can achieve breakdown points as high as 50%. Their approach is to (Simpson et al. [1992, p.440, words in parentheses added])

> start with a (preliminary) high breakdown point (HBP) estimator and perform (some) iteration(s) of a Newton-Raphson-type algorithm towards solution of the GM estimating equations

Under certain conditions then on the weight-functions (which are explained in detail in Section 5.9), the breakdown point of the preliminary estimator carries over to the final GM estimate. Hence, this preliminary estimator (our choice of which being described in Section 5.5) should have as high a breakdown point as possible.

Within the present context, the NR-procedure is thus to be described since it will be used to calculate the GM estimator (which accordingly will have a breakdown point of 50%, see Section 5.9.1). Differentiating the LHS of (5.10) with respect to b gives as Hessian

$$H = - \sum_{i=1}^{n} x_i^T w_x(x_i)[x_i w_r(r_i) - r_i w_r'(r_i)],$$

(5.11)

where

$$w_r'(r_i) = \frac{\partial w_r(r_i)}{\partial r_i} \frac{\partial r_i}{\partial b}.$$

(5.12)

Hence, the NR-rule boils down to

$$b_{j+1} = b_j + H_j^{-1} g_j,$$

(5.13)

with[9]

$$g_j = \sum_{i=1}^{n} x_i^T w_x(x_i) r_{ij} w_r(r_{ij}).$$

(5.14)

In the remainder of this chapter, a particular GM estimator is derived and its statistical properties are discussed. This means that functional forms of both $w_r(\cdot)$ and $\rho(\cdot)$ have to be chosen, as well as a HBP preliminary estimator.

5.5 Least Median of Squares

To initiate the NR-iterations, an initial estimate of β, b_0 has to be generated. Observe that this initial estimate is also needed for determining which observations are preliminary vertical outliers. Indeed, there is a simultaneity problem in that b determines r_i, and r_i determines b (see (5.10)). Given then the preliminary estimate, b_0, first-round residuals can be computed which then also determine the first-round weights for vertical outliers. Because every subsequent NR-iteration is one step closer to the solution of (5.10), in each iteration $w_r(r_i)$ is updated according to the 'latest' estimate of β.

A natural candidate for starting the numerical approximation of (5.10) is Rousseeuw's [1984] Least Median of Squares (LMS) estimator since this estimator is relatively easy to compute and has a breakdown point of 50%.[10] In this section the LMS estimator is described, together with a concomitant estimate of the variance of the residuals. Also an explanation is given is to how we compute the LMS-estimator, since it does not have a closed form solution.

---

[9] In (5.14) $r_{i,j}$ refers to the ith residual in the jth iteration.

[10] A major setback of Rousseeuw's [1984] LMS is its low efficiency, a deficit that also will be illustrated in Section 5.9.3.

5.5.1 Definition

Recall that the classic least squares technique has as objective to

$$\min_{b} \left\{ \sum_{i=1}^{n} (y_i - x_i b)^2 \right\}. \tag{5.15}$$

Replacing the summation in (5.15) by the median over all squares yields the LMS estimator

$$\min_{b} \left\{ med(y_i - x_i b)^2 \right\}. \tag{5.16}$$

Rousseeuw [1984] shows that for n observations and p explanatory variables the breakdown point of the LMS estimator is $([n/2] - p + 2)/n$, which is as high as 50% when n goes to infinity (the notation $[r]$ stands for the largest integer less than or equal to r).

5.5.2 Variance estimate

The estimated variance associated with the LMS technique is the result of a two stage procedure (see Rousseeuw and Leroy [1987]). First, an initial estimate is made according to

$$s_0 = [1 + 5/(n-p)] \sqrt{med[r_i(b_{LMS})^2]} / 0.6745, \tag{5.17}$$

where $r_i(b_{LMS})$ is residual i based on the LMS estimate b_{LMS}. Given s_0 weights are determined according to

$$w(r_i) = \begin{cases} 1 & if \ |r_i/s^0| \le 2.5 \\ 0 & otherwise \end{cases}. \tag{5.18}$$

The second step yields the estimate of the standard deviation

$$s_{LMS} = \sqrt{ \sum_{i=1}^{n} [w(r_i)r_i]^2 \Big/ \sum_{i=1}^{n} w(r_i) - p }. \tag{5.19}$$

Note that it is this standard deviation estimate we use to scale (5.1) so that the solution to (5.10) becomes invariant with respect to rescalings of the y-axis.

5.5.3 Computation

Observe that (5.16) does not have a closed form solution. To approximate the LMS estimator a *resampling* algorithm is used. In particular, the OLS estimate of some randomly drawn sample J of size $p+1$ is computed (this results a perfect fit). Given this estimate, b_{LMS}^J, the objective value, $med(y_i - x_i b_{LMS}^J)^2$, is determined and we proceed with the next drawing. The estimate with the lowest objective value, b_{LMS}, is the first approximation of the LMS estimate. We then proceed by calculating a refinement for the intercept estimate. In particular, the final estimate of the intercept is given by

$$b_{LMS_1} = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^* b_{LMS}^*),$$

(5.20)

where $x_i^* = \{x_{i,2},...,x_{i,p}\}$ and $b_{LMS}^* = \{b_{LMS_2},...,b_{LMS_p}\}^T$ (see Rousseeuw and Leroy [1987]) [11]

If all distinct samples were to be enumerated, there were far too many estimates to be made. Rousseeuw and Leroy [1987] show that the probability of getting a sub sample consisting of p non-outlying observations when the fraction of contamination equals δ, is

$$1 - [1 - (1-\delta)^p]^m,$$

(5.21)

where m is the required number of independent sub samples. Rewriting (5.18) gives the number of drawings needed to get a 'good' sub sample with probability Λ

$$m = \ln(1-\Lambda)/\ln[1-(1-\delta)^p].$$

(5.22)

In our GAUSS program Λ is set equal to 0.99, δ is set equal to 0.5, and then ten times as much trials as prescribed by (5.22) are carried out.

5.6 Down-weighting leverage points

Recall that the identification of leverage points is independent of the fitted relation (see Section 5.2.2). Rather, each of the independent observations is to be assessed according to its relative (multi-dimensional) position in the design matrix X. A classical measure of distance is the diagonal of the hat matrix (see e.g. Judge et al. [1988, p.892])

$$h_i = x_i(X^T X)^{-1} x_i^T.$$

(5.23)

This measure can be interpreted as the extent to which an observation deviates from the sample average. Related to (5.23) is the Mahalanobis distance [12]

---

[11] For a refined algorithm when different intercepts are involved see Rousseeuw and Wagner [1994].

[12] Observe that $h_i = MD_i^2/(n-1) + 1/n$ (see e.g. Rousseeuw and Van Zomeren [1990]).

$$MD(x_i) = \sqrt{[x_i - m(X)]C(X)^{-1}[x_i - m(X)]^T} , \tag{5.24}$$

where $m(\cdot)$ is the arithmetic mean, and where $C(\cdot)$ is the sample covariance matrix. Observe that the Mahalanobis distance is often referred to as the z-score.

Given the Mahalanobis distances, a weight-function $w_x(\cdot)$ can be designed. For the results of Simpson et al. [1992] to apply, this function should be of the form

$$w_x(x_i) = \min\left\{1, \left[\frac{\gamma}{MD(x_i)^2}\right]^{\alpha/2}\right\}. \tag{5.25}$$

Since the squared Mahalanobis distances follow a χ^2-distribution with p degrees of freedom, a common choice for γ is some percentile of this distribution. In that case, any observation with a concomitant Mahalanobis distance exceeding the chosen percentile is down-weighted. Further, observe that the larger is α, the more is a leverage points down-weighted. Indeed, in the extreme case of $\alpha = \infty$, any observation with a Mahalanobis distance exceeding γ is deleted, a way of 'cleaning' data which undoubtedly erases valuable information (see Section 5.2.3). In the other extreme case, $\alpha = 0$, leverage points are completely ignored and (5.10) boils down to the first order condition for a M estimator of β in (5.1). We follow traditional practise and set α equal to 1.

However, as shown by Rousseeuw and Van Zomeren [1990], the Mahalanobis distance suffers from a *masking effect*. That is, mild leverage points remain undetected if severe leverage points corrupt the data, since the latter disturb both $m(\cdot)$ and $C(\cdot)$ such that the former remain unnoticed (for a prominent exposition of this masking effect see Rousseeuw and Van Zomeren [1990]). Therefore, using (5.24) for down-weighting leverage points could be misleading. To overcome this problem robust estimates of the Mahalanobis distances, RD_i, have to be generated. For that purpose we use Minimum Volume Ellipsoid (MVE) estimates of location and scatter (see Rousseeuw [1985]). Indeed, the breakdown point of the MVE estimator is $([n/2] - p + 1)/n$, which is 50% as the number of observations goes to infinity (see Rousseeuw and Leroy [1987, p.259]). Hence, each independent observation receives a weight given by

$$w_x(x_i) = \min\left\{1, \frac{\sqrt{\chi^2_{0.975}(p)}}{RD(x_i)}\right\}. \tag{5.26}$$

Using MVE estimates of location and scale guarantees that both mild and severe leverage points are identified and down-weighted accordingly. In Figure 5.3 weight-function (5.26) is illustrated.

Before we proceed with explaining how vertical outliers are down-weighted, the MVE estimator will be described, together with some computational problems associated with it.

$$w_x(x_i)$$
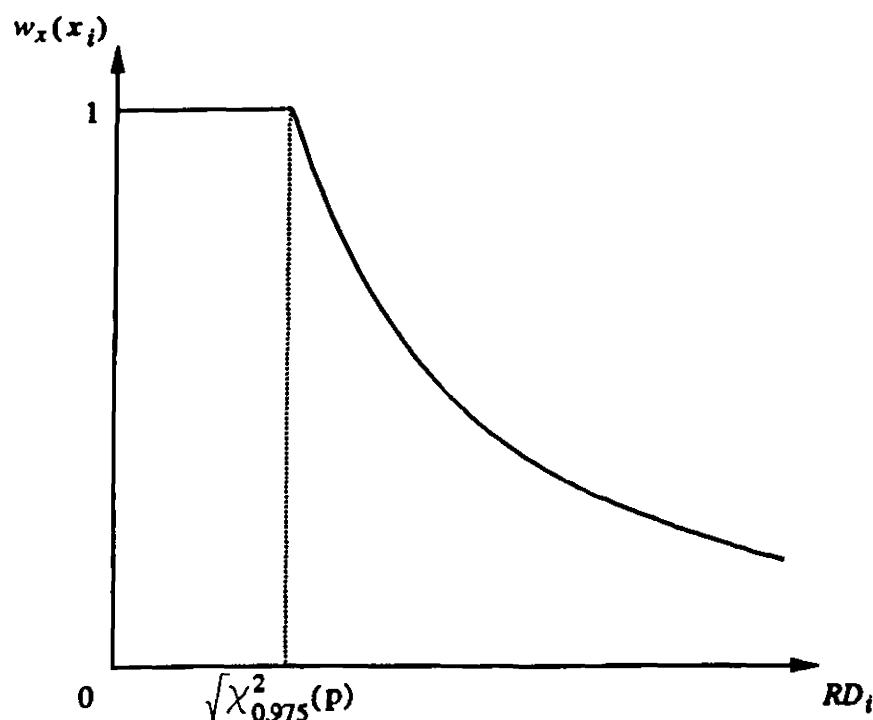
Figure 5.3 Down-weighting leverage points

5.7 Minimum Volume Ellipsoid estimation[13]

The MVE estimator of location and scale is based on the hyper ellipsoid of minimum hyper area containing at least half of the observations. The estimate of location corresponds to the centre of this hyper ellipsoid while the corresponding covariance estimate is the hyper ellipsoid multiplied by some factor to obtain consistency.

Recall that the determinant of a scatter matrix (that is, a positive definite symmetric matrix) is proportional to the squared volume of the corresponding tolerance ellipsoid. Accordingly, the MVE estimator is defined as the pair (m,C) which

minimizes $\det(C)$
(m,C)

subject to   (5.27)

$$\#\{i;\ [x_i-m]C^{-1}[x_i-m]^T < \delta^2\} \geq h,$$

where $h = [(n+p+1)/2]$. If it is assumed that the majority of the data comes from a normal distribution, δ^2 is set equal to the 50th percentile of the $\chi^2(p)$ distribution. We then denote

---

[13] This section follows roughly Section 3 of Hinloopen and Wagenvoort [1997].

$$RD(x_i) = \sqrt{[x_i - m]C^{-1}[x_i - m]^T} \; , \tag{5.28}$$

as the robust distance of case x_i.

5.7.1 A correction factor

Observe that m and C are defined such that the h-th order statistic of the robust distances is

$$\{RD(x_i)\}_h = \sqrt{\chi^2_{0.50}(p)} \; . \tag{5.29}$$

The MVE estimator does not have a closed form solution. Hence we numerically approximate the solution to (5.27) and must make sure that the results of the algorithms used to compute the MVE also satisfy (5.29). The two algorithms reported in the literature, the *resampling* algorithm and the *projection* algorithm, do not yield approximations satisfying (5.29), as will become clear below. We therefore propose to correct the *approximated* MVE-distances, $R\tilde{D}_i(\cdot)$, computed by either algorithm, by multiplying each computed robust distance with the correction factor[14]

$$\frac{\sqrt{\chi^2_{0.50}(p)}}{\{R\tilde{D}(x_i)\}_h} \; . \tag{5.30}$$

In this way the median (or more precisely, the h-th order statistic) of the $R\tilde{D}_i(\cdot)$ is always equal to the root of the 50th percentile of the χ^2-distribution with p degrees of freedom. In (5.30) the h-th percentile is chosen because the breakdown point of the MVE estimator is 50%.

To assess correction factor (5.30) we have run some simulations, the results of which are presented below. First however the two algorithms used to approximate the MVE-distances are to be described.

5.7.2 The resampling algorithm

The oldest algorithm to approximate the MVE is the *resampling* algorithm proposed by Rousseeuw and Leroy [1987, p.259-60]; see also Rousseeuw and Van Zomeren [1990, 1991]. The idea is to compute a sufficient number of ellipsoids containing half of all observations. Among these, the one with lowest volume is taken as the approximation of (5.27).

---

[14] This factor is originally conceived by Teun Kloek (personal communication).

In particular, let

$$m_K = \frac{1}{p+1}\sum_{i \in K} x_i \tag{5.31}$$

and

$$C_K = \frac{1}{p}\sum_{i \in K} (x_i - m_K)^T (x_i - m_K) \tag{5.32}$$

be respectively the arithmetic mean and corresponding covariance matrix of a subsample of $(p+1)$ different observations, indexed by $K = \{i_1,...,i_{p+1}\}$. To assure that exactly h observations are contained in the corresponding ellipsoid, C_K should be inflated or deflated by

$$f_K^2 = \underset{i=1,...,n}{med} \left\{ (x_i - m_K) C_K^{-1} (x_i - m_K)^T \right\}. \tag{5.33}$$

Indeed, the ellipsoid containing h observations corresponding to subsample K is given by $f_K^2 C_K$, with a concomitant volume proportional to

$$\det(f_K^2 C_K). \tag{5.34}$$

The algorithm consists of computing the value of (5.34) for many different subsamples, after which the subsample $\hat{K}$, corresponding to the lowest value of the objective function, is retained. The MVE-approximation of location and scatter are then respectively given by

$$m_{res} = m_{\hat{K}} \tag{5.35}$$

and

$$C_{res} = \frac{f_{\hat{K}}^2}{\chi_{0.50}^2(p)} C_{\hat{K}}. \tag{5.36}$$

Robust distances are then obtained by inserting (5.35) and (5.36) into (5.28). Observe that (5.29) holds if m_{res} and C_{res} are based on the true ellipsoid of minimum volume.

For most problems of interest not all possible subsamples can be in practise considered.[15] Hence, the *actual* ellipsoid of minimum volume, corresponding to (m^*, C^*), will probably not be found. Therefore, $(m_{res}, C_{res}) \neq (m^*, C^*)$ and (5.29) does not necessarily hold. Applying then our correction factor (5.30) to the approximated MVE-distances ensures that (5.29) is valid.

---

[15] For an indication as to how many subsamples are to be considered to adequately approximate the actual ellipsoid of minimum volume, see Rousseeuw and Leroy [1987, p.260] (the probabilistic rule is very similar to that concerning the adequate number of drawings needed to compute the LMS estimator). Again we perform 10 times as many drawings as prescribed for the worst possible data corruption (50%) under a probability of 99% for drawing a 'clean' subsample. See also Cook and Hawkins [1990].

An improvement to the resampling algorithm as such is due to Rousseeuw and Van Zomeren [1991]. They determine an empirical factor c, which depends both on n and p, such that on average the 97.5th percentile of the $RD(x_i)$ approaches the square root of the χ^2-distribution with the appropriate degrees of freedom. In particular, (5.36) is multiplied by

$$c(n,p)^2 = [1 + 15/(n-p)]^2. \tag{5.37}$$

Our simulations show that (5.37) indeed improves the approximation to the final estimate of the MVE (see Section 5.7.4) but does not completely restore the validness of (5.29).

Another improvement was already proposed by Rousseeuw and Leroy [1987, p.260]. They assign to each observation a weight $w(\cdot)$ according to

$$w(x_i) = \begin{cases} 1 & \textit{if } (x_i - m_{res})C_{res}^{-1}(x_i - x_{res})^T \leq \kappa \\ 0 & \textit{otherwise} \end{cases}, \tag{5.38}$$

where, for instance, $\kappa = \chi^2_{0.975}(p)$. More efficient (re-weighted) estimators for location and scatter are then suggested as being respectively

$$m'_{res} = \sum_{i=1}^{n} w(x_i)x_i \bigg/ \sum_{i=1}^{n} w(x_i) \tag{5.39}$$

and

$$C'_{res} = \frac{\sum_{i=1}^{n} w(x_i)(x_i - m'_{res})^T(x_i - m'_{res})}{\sum_{i=1}^{n} w(x_i) - 1}. \tag{5.40}$$

In Subsection 5.7.4 we test both our correction factor and the one step improvement of Rousseeuw and Leroy, as well as the situation in which we apply correction factor (5.30) on robust distances which are based on (5.39) and (5.40). We always multiply (5.36) by (5.37). Indeed, under all different versions of the resampling algorithm most unsatisfactory results are obtained when this initial correction is not made. The simulation results presented in Subsection 5.7.4 show that the procedure in which correction factor (5.30) is applied to robust distances obtained from (5.39) and (5.40) results in the most satisfactory approximations of the distances (5.28) based on the MVE estimates of location and scale.

5.7.3 The projection algorithm

We also consider a second algorithm for the MVE, the *projection* algorithm of Rousseeuw and Van Zomeren [1990], which is based on Gasko and Donoho [1982]. This algorithm is built around

$$u_i = \max_{v} \frac{|x_i v^T - L(x_1 v^T, \dots, x_n v^T)|}{S(x_1 v^T, \dots, x_n v^T)}, \tag{5.41}$$

which is the exact one-dimensional version of (5.24) applied to the projections $x_i v^T$ if $L(\cdot)$ and $S(\cdot)$ are the MVE-estimates of location and scatter respectively. The latter are defined as

$$L(z_1, \dots, z_n) = (z_j + z_{j-h+1})/2 \tag{5.42}$$

and

$$S(z_1, \dots, z_n) = \gamma[z_j - z_{j-h+1}] \tag{5.43}$$

where $z_1 \le z_2 \le \dots \le z_n$ is any set of numbers, where $z_j - z_{j-h+1}$ is the smallest of the differences

$$z_h - z_1, z_{h+1} - z_2, \dots, z_n - z_{n-h+1} \tag{5.44}$$

and where γ is some correction factor to make $S(\cdot)$ an approximately unbiased (or at least consistent) estimator of the unknown scale.

In principle all directions v should be considered to compute (5.27). Because in practise this is not possible, Rousseeuw and Van Zomeren [1990] propose to restrict the set of directions to all v defined as $x_l - M$, $l = 1, \dots, n$, with $M = \{ med(x_{j1}), \dots, med(x_{jp}) \}$.

Observe however that the u_i in (5.41) do not yet satisfy (5.29). Indeed, in the special case of one dimension we have

$$u_i = \frac{|x_i - (x_j + x_{j-h+1})/2|}{x_j - x_{j-h+1}} \tag{5.45}$$

where $x_1 \le x_2 \le \dots \le x_n$. In this case we can compute the h-th order statistic of the u_i exactly. As a first step, we find the remarkable identity

$$\{|x_i - m|\}_h = \max_{j-1+h \le i \le j} |x_i - m| = (x_j - x_{j-h+1})/2, \tag{5.46}$$

because h of the x_i's are in between x_{j-h+1} and x_j. Hence,

$$\{u_i\}_h = \left\{ \frac{|x_i - m|}{S(\cdot)} \right\}_h = \frac{1}{2\gamma}. \tag{5.47}$$

In order to obtain robust distances $RD(x_i)$ satisfying (5.29) we need to put

$$\gamma = \frac{1}{2\sqrt{\chi^2_{0.50}(1)}}.$$

(5.48)

In general dimensions we can achieve this by multiplying the u_i's by correction factor (5.30). Robust distances based on the projection algorithm are then given by

$$RD(x_i) = \frac{\sqrt{\chi^2_{0.50}(p)}}{\{u_i\}_h} u_i.$$

(5.49)

5.7.4 Testing the MVE-approximations

To assess the empirical performance of correction factor (5.30) we have tested different versions of both the resampling and projection algorithm to approximate (5.27). Table 5.1 reports our simulation results. Each experiment consisted of generating a standard normal random variable for which the MVE-distances were approximated, using various versions of the resampling algorithm (lines I-V) and the projection algorithm (lines VI-VII). It was then examined how many of the approximated squared MVE-distances exceeded the 97.5th percentile of the $\chi^2(p)$-distribution, $\hat{\alpha}$. This was repeated 10,000 times both for a small sample ($n = 50$) and for a large sample ($n = 250$), as well as for a number of explanatory variables. In Table 5.1 the mean outcomes of $1 - \hat{\alpha}$ are reported.

The rows labelled (I) in Table 5.1 contain the results obtained when using the resampling algorithm without any correction. Clearly, (5.29) is not satisfied. The impact of the initial correction factor (5.37) on the resampling algorithm is revealed in the second row. Indeed, a significant improvement is observed although apart from the one dimensional case the MVE-distances are not adequately approximated. However, in all other versions of the resampling algorithm (lines III through V) we have used (5.37). Observe also that the initial correction factor has much more impact in the small sample case ($n = 50$) than in case of the large sample ($n = 250$). This is due to correction factor (5.37) approaching 1 if n increases. The third row entries (III) involve the resampling algorithm with the one-step improvement as suggested by Rousseeuw and Leroy [1987, p.260]. Some improvement occurs, although in all cases considered the 97.5th percentile is underestimated. Algorithm IV consists of applying correction factor (5.30) to (II). This however does not give any satisfactory result. Both the one-step improvement of Rousseeuw and Leroy [1987] and correction factor (5.30) are used in the fifth algorithm. This version of the resampling algorithm has the best performance of all versions considered. Indeed, since the third algorithm consistently underestimates the 97.5th percentile there is still room for additional refinement, which we achieve by applying (5.30) to this algorithm.

The projection algorithm (VI) almost never yields appropriate distance approximations (of course, for $p = 1$ this follows immediately from the discussion above). Applying correction factor (5.30) to this algorithm does improve its performance, especially when the sample is large, as can be seen from the rows marked VII.

Table 5.1 Simulated 97.5th percentiles using resampling and projection algorithms[a]

Algorithm[b]	n =50				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
I	0.936926	0.878970	0.854576	0.839810	0.829328
II	0.982440	0.957838	0.949937	0.946205	0.944832
III	0.968137	0.956340	0.953417	0.951283	0.950901
IV	0.937292	0.879456	0.854413	0.838526	0.828575
V	0.968073	0.964506	0.962086	0.959474	0.958450
VI	0.985327	0.974833	0.987927	0.992823	0.997894
VII	0.937170	0.975820	0.967289	0.955688	0.945347
	n =250				
I	0.963207	0.951092	0.942612	0.932714	0.922099
II	0.973080	0.964387	0.959334	0.952085	0.944556
III	0.961111	0.960795	0.961163	0.960228	0.959358
IV	0.963270	0.950656	0.942743	0.932340	0.921929
V	0.972621	0.972188	0.971987	0.971478	0.971067
VI	0.994732	0.994344	0.996856	0.998490	0.999479
VII	0.962763	0.985931	0.981385	0.974881	0.967968

[a] Each cell contains the mean of the estimated percentile over the 10,000 runs.
[b] I   Resampling
    II   Resampling with correction factor (5.37), denoted by resampling*
    III   Resampling* with Rousseeuw and Leroy's one step improvement
    IV   Resampling* with correction factor (5.30)
    V   Resampling* with Rousseeuw and Leroy's one step improvement *and* correction factor (5.30)
    VI   Projection
    VII Projection with correction factor (5.30)

To conclude, in all cases considered the resampling algorithm with Rousseeuw and Van Zomeren's correction factor (5.37), Rousseeuw and Leroy's one step improvement *and* correction factor (5.30) yields reliable approximations of the MVE distances. Further, if the sample is large, applying correction factor (5.30) to the projection algorithm also results in appropriate distance approximations. For the simulations concerning efficiency of the GM estimator (reported in Subsection 5.9.3) we have used algorithm V.

5.8 Down-weighting vertical outliers

Having decided on the preliminary estimator, and determined how to down-weight leverage points, the only thing left to do in order to complete the construction of the GM estimator is to design weights for vertical outliers. Recall from the first order condition defining a GM estimator (see Section 5.4) that vertical outliers are down-weighted according to a weight-function of the form

$$w_r(r_i) = \frac{\partial \rho(r_i)}{\partial r_i} \bigg/ r_i = \psi(r_i)/r_i. \tag{5.50}$$

In this section an appropriate choice of $\psi(\cdot)$, the *score*-function, will be made.

Observe that the *score*-function is proportional to the rate of change in the objective (5.10) due to an infinitely small change in r_i. In particular, it (indirectly) measures the influence of an infinitesimal change in r_i on the value of the statistic implicitly defined by (5.10). Indeed, there is a one-to-one relation between the score-function and the influence function.[16] Hence, to examine the influence of vertical outliers on objective (5.10), and to construct $w_r(r_i)$ accordingly, we only have to consider $\psi(r_i)$.

In Figure 5.4 two possible candidates for $\psi(\cdot)$ are depicted (Huber's [1964] proposal and Tukey's bi-square function; see Beaton and Tukey [1974]), together with OLS' score-function. To begin with the latter, it is immediate that the influence of an observation could be unbounded. As shown in example 5.1, the breakdown point of OLS is 0%: a large enough vertical outlier can cause this technique to produce any estimate. Huber's [1964] proposal for $\psi(\cdot)$ bounds the influence of vertical outliers. Any observation with an absolute (scaled) residual exceeding b is downweighted. However, this rule does not distinguish between mild and severe vertical outliers. Score-functions which are redescending on the other hand, such as the depicted bi-square function of Tukey, do make this distinction. In particular, observations with low residuals are fully taken into account, whereas data with residuals above a certain threshold value are down-weighted or completely ignored. The latter happens if the concomitant residual exceeds the *rejection point* (denoted by c in Figure 5.5; ignore for the moment point a).

We base weight-function (5.50) on Tukey's bi-square function. This function takes the form

$$\psi(r_i) = \begin{cases} r_i[1 - (r_i/\alpha)^2]^2, & |r_i| < \alpha, \\ 0, & |r_i| \geq \alpha. \end{cases} \tag{5.51}$$

The exact value of the tuning constant α is a compromise between efficiency and robustness. The lower is α, the more are vertical outliers down-weighted, and the more robust the concomitant estimator. In the limit, when α is 0, all observations are rejected. On the other hand, if α goes to infinity, no observation is down-weighted and the estimator is as vulnerable to vertical outliers as OLS. However, down-weighting does not come without a loss. It will increase the variance of the estimator

---

[16] In case of M estimators the influence function is proportional to $\psi(\cdot)$, whereas the weight-function based on leverage points enters in addition the influence function of GM estimators.
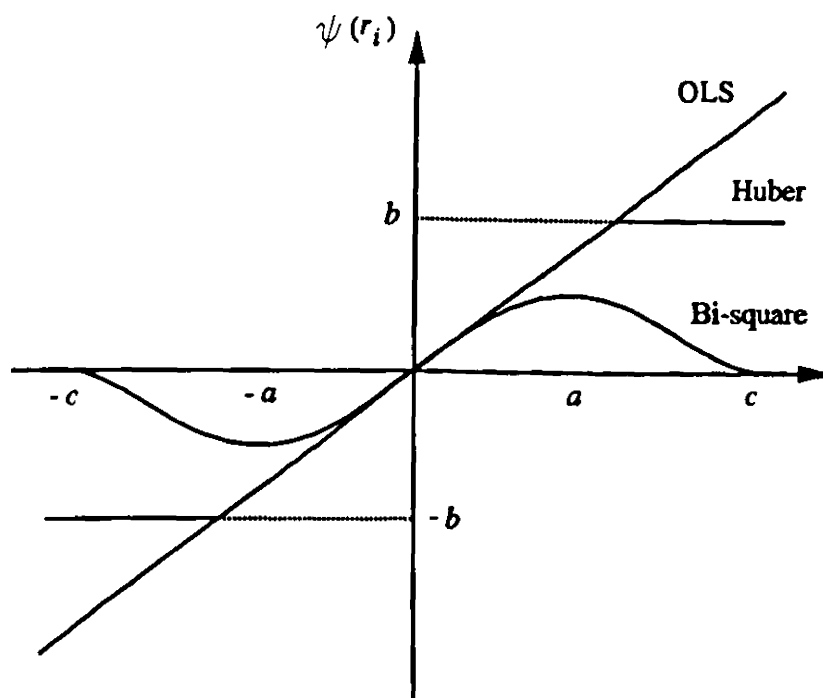
Figure 5.4 *Score*-functions

and thus lowers the estimator's efficiency. It is common practise to settle for an efficiency level of 95%. If the residuals are assumed to follow a standard normal distribution (as we obtain by scaling model (5.1) with (5.19)), it follows that α should be set at 4.685 to obtain this efficiency level (in Section 5.9.3, simulation results as to the validness of this value will be presented).

(5.50) now becomes

$$w_r(r_i) = \begin{cases} 1, & r_i = 0, \\ [1-(r_i/4.685)^2]^2, & 0 < |r_i| < 4.685, \\ 0, & 4.685 \le |r_i|. \end{cases} \tag{5.52}$$

In Figure 5.5, weight-function (5.52) is depicted together with a standard normal distribution. Observe that (in absolute terms) large residuals are down-weighted, while only extreme vertical outliers (that is, those observations with an absolute concomitant residual above 4.685) are ignored.

5.9 Statistical properties of the multi-step GM estimator

The GM estimator to be used in the next chapter is now described in full detail. Before applying it however, its statistical properties should be examined. In this section, the estimator's breakdown point, its asymptotic behaviour and its relative efficiency are discussed.
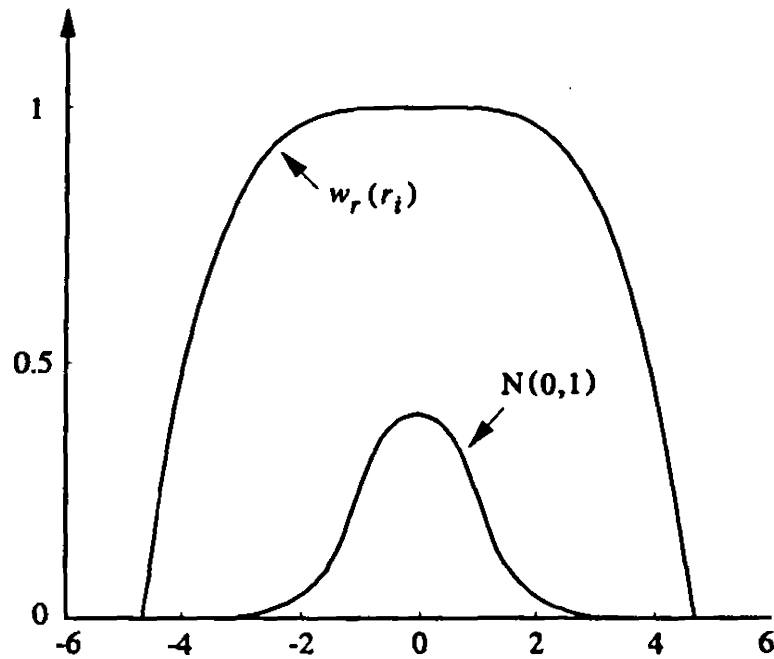
Figure 5.5   **The weight-function for vertical outliers based on Tukey's bi-square function in relation to a Standard Normal distribution**

5.9.1 Breakdown point

To derive the estimator's breakdown point, assume that in some dataset $\{y,X\}$ the first $n-l$ observations are free from outliers, while the remaining l observations can take on any value, with $l \leq (n-2)/2$ and $n \geq l+p$. As to the breakdown point of a multi-step GM estimator based on a nondecreasing score-function, applied to this type of data, Simpson et al. [1992] prove the following.

Theorem 5.1 (Simpson et al. [1992])

$$\varepsilon^*(b_{GM}, \{y,X\}) = \frac{l}{n},$$

if the following three conditions hold

(i) $\psi(r)/r \geq \alpha_0 > 0,$ $|r| \leq a,$

(ii) $\partial\psi(r)/\partial r \geq \alpha_1 > 0,$ $|r| \leq a,$

with $a > \kappa$, and

(iii) any set of $n/2 - l$ 'good' points has a linearly independent subset of size p.

Verification

With $\psi(\cdot)$ chosen as in (5.51), the expressions of interest are

$$\frac{\psi(r_i)}{r_i} = \begin{cases} \left[1-(r_i/\alpha)^2\right]^2, & |r_i| < \alpha, \\ 0, & |r_i| \geq \alpha, \end{cases} \tag{5.53}$$

and

$$\frac{\partial \psi(r_i)}{\partial r_i} = \begin{cases} \left[1-(r_i/\alpha)^2\right]\left[1-5(r_i/\alpha)^2\right], & |r_i| < \alpha, \\ 0, & |r_i| \geq \alpha. \end{cases} \tag{5.54}$$

If $\alpha = 4.685$ the first two conditions of Theorem 5.1 hold for $a = 2.095$. Observe that the tuning constant κ is the one used when estimating the initial variance. Hence, in our case it equals 0.6745 (see (5.17)) and is thus below a. The third condition is always assumed to hold, since if not, the LMS estimator (and hence the GM estimator) is undetermined. To put it differently, if condition (*iii*) is not fulfilled, more than 50% of the linearly independent data points are corrupted and any estimate of β is arbitrary.

As mentioned, Theorem 5.1 is valid for nondecreasing $\psi(\cdot)$ functions. However, Tucky's bi-square function is redescending, and the reader may conjecture that Theorem 5.1 does not hold for the GM estimator derived in this chapter (observe that Theorem 5.1 is applicable if Huber's [1964] proposal for $\psi(\cdot)$ is used to construct $w_r(\cdot)$, a score-function according to which vertical outliers are down-weighted less strongly than according to Tucky's bi-square function). But let us reason why Theorem 5.1 also applies in case of redescending score-functions.

The reason as to why the GM estimator could break down is that the NR-algorithm may not converge. This happens when in some iteration the Hessian is not invertible. Observe that the Hessian is given by (see (5.11) and (5.50))

$$H = \sum_{i=1}^{n} x_i^T w_x(x_i) \frac{\partial \psi(r_i)}{\partial r_i} x_i, \tag{5.55}$$

and that it is positive definite whenever all its diagonal entries are greater than zero. For redescending score-functions this is not guaranteed, since in this case any diagonal entry in (5.55) could go negative (because $\partial \psi(r)/\partial r$ could be negative).[17] However, as long as the NR-algorithm is initiated 'close' enough to the solution of (5.10), this is precluded and convergence will always occur (remember that

---

[17] Simpson et al. [1992, Remark 2.1] conjecture that in this case p data points can be manipulated such that the resulting Hessian is positive definite. Hence, also for GM estimators based on redescending $\psi(\cdot)$ functions the breakdown point is $1/m$. But this conjecture is hard to maintain, since manipulating data changes the problem and will lead to different estimates.

the NR-method is a local algorithm). We presume that because of the HBP of the LMS estimator, it is always possible to start the NR-procedure close enough to the solution of (5.10) for convergence to take place. It thus follows that Theorem 5.1 also holds for redescending score-functions. Accordingly, our GM estimator has a breakdown point of $l/n = 1/2 - 1/n$, which goes to 50% as n increases.

5.9.2 Consistency and asymptotic normality

To examine the large sample properties of the GM estimator, first its covariance matrix is to be derived. For that purpose, consider a Taylor expansion around $b_{GM} = \beta$ for the first order condition (5.10)

$$0 = \sum_{i=1}^{n} x_i^T w_x(x_i) \psi(r_i) \approx \sum_{i=1}^{n} x_i^T w_x(x_i) \psi(e_i) - \sum_{i=1}^{n} x_i^T w_x(x_i) \frac{\partial \psi(e_i)}{\partial e_i} x_i [b_{GM} - \beta], \tag{5.56}$$

where second and higher order terms are ignored. Performing obvious manipulations results

$$\sqrt{n} [b_{GM} - \beta] \approx \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^T w_x(x_i) \frac{\partial \psi(e_i)}{\partial e_i} x_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^T w_x(x_i) \psi(e_i) \right). \tag{5.57}$$

Hence, in large samples, the covariance matrix of b_{GM} approximately equals

$$\Sigma_{b_{GM}} = \left(\sum_{i=1}^{n} x_i^T w_x(x_i) \frac{\partial \psi(r_i)}{\partial r_i} x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i^T w_x(x_i)^2 \psi(r_i)^2 x_i \right) \left(\sum_{i=1}^{n} x_i^T w_x(x_i) \frac{\partial \psi(r_i)}{\partial r_i} x_i \right)^{-1}. \tag{5.58}$$

Observe that with (5.58) concomitant t-values can be computed.

If the multi-step GM estimator is based on MVE estimates of location and scale, and if leverage points are down-weighted according to some function of the form given in (5.26), Simpson et al. [1992] show that the following holds

Theorem 5.2 (Simpson et al. [1992])

$$b_{GM} \xrightarrow{d} N(\beta, \Sigma_{b_{GM}})$$

if the following four conditions hold

(*i*) $\psi(\cdot)$ is bounded and continuous,

$$(ii) \quad \left|\frac{\partial \psi(r_i)}{\partial r_i}\right|, \left|\frac{r_i \partial \psi(r_i)}{\partial r_i}\right|, \left|\frac{\partial^2 \psi(r_i)}{\partial r_i^2}\right|, \left|\frac{r_i \partial^2 \psi(r_i)}{\partial r_i^2}\right|, \left|\frac{r_i^2 \partial^2(r_i)}{\partial r_i^2}\right| < \infty.$$

$$(iii) \quad \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i^T w_x(x_i)^2 E\left[\psi(r_i)^2\right] x_i = A, \text{ and } \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i^T w_x(x_i) E[\partial \psi(r_i)/\partial r_i] x_i = B.$$

$$(iv) \quad E[\psi(r_i v)] = 0, \text{ and } E\left[r_i v \frac{\partial \psi(r_i v)}{\partial r_i}\right] = 0,$$

where $|\cdot|$ is the supremum norm, A and B some symmetric positive definite matrices, $E(\cdot)$ the expectations operator, and v some scalar.

Verification

The first condition follows immediately from (5.51). Condition (*ii*) is readily verified to hold given (5.54) and

$$\frac{\partial^2 \psi(r_i)}{\partial r_i^2} = \begin{cases} -4r_i\left[3 - 5(r_i/\alpha)^2\right]/\alpha^2, & |r_i| < \alpha, \\ 0, & |r_i| \geq \alpha. \end{cases} \tag{5.59}$$

Assuming the third condition to be met is equivalent to assuming that the Hessian is positive definite (and given by some matrix B; see also the discussion in Section 5.9.1), an assumption which is harmless to make. Finally, condition (*iv*) follows immediately if the error terms are assumed (as we do) to be independent and identically distributed according to a (standard) normal distribution.

It thus follows that the GM estimator is asymptotically normal distributed with mean β, and covariance matrix $\Sigma_{b_{GM}}$. Observe that Theorem 5.2 also implies that the GM estimator is consistent, since the estimator converges to the mean of its distribution, being the true value of the unknown in (5.1).

5.9.3 Efficiency[18]

Finally, to examine the efficiency of the multi-step GM estimator compared to OLS and LMS under different types and fractions of data contamination, we have performed a number of simulations. A single experiment began with generating a matrix of explanatory variables, X, of length 200, consisting of an intercept and one explanatory variable drawn from a standard normal distribution.

---

[18] This subsection is based on Section 4 of Hinloopen and Wagenvoort [1997].

Table 5.2 Simulation results with uncontaminated data,
$\beta_1 = \beta_2 = 1$, n=200[a]

	OLS	LMS	GM
β_1	0.9989610	1.0023398	0.9991616
	0.0049485	0.0421095	0.0052413
	0.0049474	0.0421040	0.0052406
β_2	1.0012882	1.0009535	1.0011469
	0.0050393	0.0378368	0.0053645
	0.0050376	0.0378358	0.0053632

[a] Each cell contains the mean of the estimated parameter value, and
the mean squared error and the variance over the 10,000 runs.

Then a response variable, y, was created according to $y = X\beta + e$ with β_1 and β_2 set equal to unity and where e_i is *idd* $N(0,1)$, $i = 1,...,200$. Given these data we re-estimated β using the OLS, LMS and GM estimators, the results of which are presented in Table 5.2.

Examining Table 5.2 reveals that OLS both yields the lowest variance and mean squared error. This is, of course, no surprise since under the data generating process outlined above OLS is the Uniformly Minimum Variance Unbiased Estimator. Note however that both the variance and the mean squared error of the GM estimator are just a little higher. Indeed, the efficiency of the GM estimator is 94.4% for the intercept and 93.9% for the slope parameter.[19] Rousseeuw's LMS estimator on the other hand performs less well in terms of efficiency. For the intercept and slope parameter it appears to be a mere 11.8% and 13.3% respectively. This is considered a setback of the LMS estimator, despite its appealing HBP. Indeed, it is the efficiency gain that makes the GM estimator superior over LMS. Note, however, that we need the LMS in order to compute the GM estimator in the first place.

Next we successively corrupted the explanatory variables, the response variable and both the independent and dependent variables by replacing (randomly drawn) $s\%$ of the observations by random values drawn from a normal distribution with zero mean and variance 100. With these polluted data we re-estimated β using all three estimators. This process was repeated 10,000 times for s equal to 5, 10, 15, 20 and 25, the results of which are summarized in Table 5.3. Indeed, we could go up to corrupt 50% of the data considering the breakdown-point of both the LMS and GM estimator. However, as discussed in Section 5.2.1, routine data are thought to contain 1%-10% gross errors. Clearly, Table 5.3 includes all interesting percentages of data contamination.

From Table 5.3 we can draw the following conclusions. First, looking at the variance over the 10,000 runs we see that if only leverage points are constructed OLS is most efficient if 10% or more data contamination is involved (in case of only 5% data pollution by leverage points the GM estimator is most efficient). This is however little consolation considering the bias in estimates this technique yields (even if only 5% of the data are replaced by leverage points OLS returns very unreliable

[19] That the efficiency of the GM estimator is not exactly 95%, as used when determining α in (5.51), is due to the fact that now also leverage points are down-weighted.

Table 5.3 Simulation results with data contamination, $\beta_1 = \beta_2 = 1$, $n = 200$[a]

Leverage points constructed			Vertical outliers constructed			Leverage points and vertical outliers constructed		
OLS	LMS	GM	OLS	LMS	GM	OLS	LMS	GM
5% Pollution								
0.200567	0.995782	0.998774	0.949351	0.998133	0.998666	0.196364	0.998468	0.998783
0.651876	0.041557	0.005486	0.032132	0.041833	0.005632	0.731822	0.041477	0.005538
0.012784	0.041539	0.005484	0.029566	0.041829	0.005630	0.085991	0.041474	0.005536
0.199080	0.994961	0.999903	0.953168	0.998214	1.000576	0.194158	0.999168	1.000575
0.654344	0.037870	0.005787	0.033310	0.038128	0.005697	0.739740	0.038582	0.005801
0.012872	0.037845	0.005787	0.031207	0.038125	0.005697	0.090358	0.038582	0.005800
10% Pollution								
0.093830	0.991806	0.999576	0.898336	0.998771	0.998250	0.095341	0.996986	0.999447
0.823991	0.041783	0.005945	0.065004	0.041443	0.006232	0.866717	0.041690	0.005859
0.002847	0.041715	0.005945	0.054668	0.041441	0.006229	0.048310	0.041681	0.005859
0.093759	0.989552	0.999188	0.904547	0.997718	1.000113	0.095719	0.992524	1.000337
0.824204	0.038624	0.006159	0.065180	0.037663	0.006258	0.866904	0.038170	0.006165
0.002931	0.038515	0.006158	0.056068	0.037658	0.006258	0.049179	0.038114	0.006165
15% Pollution								
0.057845	0.986834	0.999977	0.848411	0.999543	0.998582	0.057884	0.993765	1.000368
0.888980	0.041591	0.006472	0.107698	0.040655	0.007167	0.920049	0.040990	0.006512
0.001323	0.041417	0.006472	0.084718	0.040655	0.007165	0.032465	0.040951	0.006512
0.056900	0.984971	0.996509	0.847506	1.002215	0.997990	0.057706	0.991309	0.998740
0.890819	0.040720	0.006712	0.110691	0.038850	0.007157	0.920038	0.038975	0.006594
0.001383	0.040494	0.006700	0.087437	0.038845	0.007152	0.032120	0.038899	0.006592
20% Pollution								
0.041390	0.980704	0.999659	0.804535	1.000447	0.998151	0.044029	0.989080	0.999559
0.919811	0.042421	0.006678	0.143066	0.041214	0.007736	0.938577	0.040697	0.006673
0.000877	0.042049	0.006678	0.104859	0.041213	0.007733	0.024696	0.040578	0.006673
0.040943	0.978557	0.995468	0.798557	0.996269	0.996417	0.041805	0.984524	0.997812
0.920667	0.040763	0.006809	0.149329	0.038140	0.007691	0.943166	0.038602	0.006821
0.000876	0.040303	0.006789	0.108750	0.038126	0.007678	0.025029	0.038362	0.006817
25% Pollution								
0.030888	0.967603	0.998027	0.747312	0.998391	0.995502	0.031209	0.981777	0.998596
0.939851	0.048125	0.006839	0.190763	0.039194	0.008250	0.959140	0.042082	0.006932
0.000674	0.047076	0.006835	0.126912	0.039191	0.008230	0.020584	0.041750	0.006930
0.030397	0.963753	0.994769	0.749810	0.999797	0.997402	0.029549	0.982081	0.997499
0.940816	0.045688	0.007214	0.197492	0.037467	0.008487	0.961735	0.040121	0.007231
0.000687	0.044374	0.007186	0.134897	0.037467	0.008480	0.019959	0.039800	0.007225

[a] Each cell contains the mean of the estimated parameter value, and the mean squared error and the variance over the 10,000 runs. The top row concerns the intercept, and the bottom row concerns the slope.

estimates). In all other cases (that is, when only vertical outliers or both leverage points and vertical outliers are constructed) the GM estimator is most efficient. Moreover, under all types and fractions of data contamination the GM estimator is more efficient than LMS.

Second, and more important in the context of data contamination, examining both efficiency and precision (that is, comparing the mean squared errors) reveals that the GM estimator is (far) superior to both OLS and LMS, given the types of outliers and amounts of data corruption considered here. Indeed, considering the performance of the GM estimator when no data anomalies are involved, it is clear that this technique yields reliable estimates for a wide range of data.

5.10 Conclusions

Any empirical dataset is likely to contain outlying observations. The researcher trying to find empirical evidence for some hypothesis either ignores these outliers, or deals with them in one way or the other. Using, in the former case, classical (OLS-like) estimation techniques can have quite dramatic consequences, since any outlying observation can induce classic estimators to yield any estimate with any level of significance. As shown in Table 5.3, the bias of OLS is already substantial if only 5% of the data are corrupted.[20]

In this chapter an example of *accommodating* outlying observations is presented. In particular, a particular General M estimator is described, and its statistical properties are derived. It is shown that the GM estimator has a breakdown point of 50%, that it is consistent, and that it is relatively efficient. In the next chapter the GM estimator is applied to existing datasets. The resulting estimates are then compared with those previously published, based on classic estimation techniques.

---

[20] Observe however that OLS is less sensitive to the construction of vertical outliers than to the construction of leverage points (this is also observed by Rousseeuw and Leroy [1987]). In our case, this is due to the way in which we have corrupted the data. With respect to the model under which the original data are generated, constructing leverage points implies that these observation also become vertically outlying. And because the original slope parameter (β_2) differs from zero, these implicit vertical outliers are much bigger than the ones which are constructed separately. Hence their stronger impact on the estimates.

6 Empirical research in Industrial Organization: a re-interpretation

6.1 Introduction

According to Martin [1993, p.447] an 'empirical renaissance in industrial economics' started somewhere around 1983, an uplift in interest for empirical research which has not yet declined.[1] This renewed interest in empirical research within the sphere of Industrial Organization is the (inevitable) reaction to the preceding decade, an era in which empirical work had fallen out of fashion.

In the 'new empirical industrial economics', three distinct trends can be identified (Bresnahan and Schmalensee [1987, p.373-4]).

> First, a large fraction of recent work employs new sources of data or data sets constructed in new ways from traditional sources...A second trend visible in recent work is the growing tendency to exploit contemporary advances in economic theory and econometric models...A third trend apparent in the recent empirical literature is a shift toward the firm, rather than the industry, as the unit of observation.

The present chapter is an example of an analysis falling within the second trend. It is concerned with the application of the GM estimator developed in Chapter 5 to existing data sets. As Bresnahan and Schmalensee [1987, p.374] remark

> empirical researchers in the field have become more willing and able to exploit the latest advances in econometric method and thus to move well beyond exclusive reliance on ordinary least squares.

While this observation may be correct, the data sets used here are taken from empirical studies, published within the last five years, all using OLS. The analysis can thus be seen as a (preliminary) test of the renaissance's second trend.

Note in passing that for the comparison of the OLS and GM regression results we first have to replicate the original studies, a task which according to Dewald et al. [1986, p.600][2]

> is an essential component of scientific methodology. Only through replication of the results of others can scientists unify the disparate findings of various researchers in a discipline into a defensible, consistent, coherent body of knowledge.

If anything, the present chapter (mildly) supports the conclusion of Dewald et al. [1986, p.587] that 'inadvertent errors in published empirical articles are a commonplace rather than a rare occurrence'[3]

---

[1] This view is shared by many others. See e.g. Bresnahan and Schmalensee [1987]

[2] Dewald et al. [1986] report on the elaborate and difficult exercise of replicating 90 empirical studies.

[3] It is therefore that well-established journals such as the *Journal of Econometrics* and *Econometrica* require authors of empirical papers to make available their data sets for possible replication.

The chapter evolves as follows. The next section contains a description of the way in which we have collected the data. After reviewing most of the received material, four studies were chosen to be re-estimated. The results of this exercise are presented in the following four sections. Two concluding remarks are stated in Section 6.7.

6.2 Collecting data

In order to illustrate the difference in estimates between the GM estimator and OLS, data sets had to be collected, preferably those for which OLS regression results are reported. For that purpose, researchers who had published empirical studies over the last one-and-a-half decade, within the sphere of Industrial Organization, were requested to make available their data.[4] In total 47 authors were asked to supply their data set.

In 17 cases there was never any reply (which is 36%, a fraction comparable with Dewald et al.'s [1986] 34%), while in three other cases reference was made to a co-author who subsequently also did not reply. Although author's coordinates were retrieved as accurately as possible, in one case the wrong person received the request.[5] The 'true' author was never to be found however. Of the 26 studies left, in two cases the data could not be provided because these were confidential,[6] in one case the data had to be bought (which we did not), and in another 3 cases the data were impossible to retrieve. This left us with a possible supply of 20 data sets, which is about 43% of the total number requested.

In two cases the data have not yet been supplied. In another case transfer was to be carried out over the Internet, a transfer that failed however.[7]

From the 17 studies left, 8 data sets were provided on floppy disc, and 9 in hard-copy. Of the latter, 7 turned out to yield quite different results than those originally reported in the literature. This could be due to a number of reasons. First, even if a scanner is used, the probability that there is an exact one-to-one relation between those observations received on paper, and those being put in the computer is low (especially handwritten (!) data are in many cases difficult to decipher). Second, the printout of data could well be (a bit) different from that being used for the final (published) analysis. All in all, in two cases we have been able to replicate exactly the results based on the data received in hard-copy. One of these studies will be discussed in detail in Section 6.5 below.

From the 8 data sets provided on floppy disc, one only consisted of a small part of the original data (and the remainder is unavailable). That leaves us with 7 ready-to-use data sets, 4 of which are still under consideration. The other three will be discussed in extensive detail in Sections 6.3, 6.4 and 6.6.

---

[4] The economics journals which were consulted, included the *International Journal of Industrial Organization*, the *Journal of Industrial Economics*, the *RAND Journal of Economics*, the *Review of Economics and Statistics*, and the *Review of Industrial Organization*.

[5] Which the alleged author acknowledged with some regret.

[6] Incidently, both these studies were based on German data.

[7] This was due to the present author's ignorance of the electronic highway.

Before we proceed some 'fair play' should be mentioned. In what follows, four studies will be critically re-examined, and in most cases the published results will be questioned. However, this whole exercise could not have been carried out if the data were not supplied in the first place. In that respect I would like to thank Professors Bloch, Connor, Mata, Mueller, Peterson, and Sial for making available their data in such an accessible manner.

6.3 National brand-private label price differences

Ever since Joe Bain published in 1951 his article 'Relation of profit rate to industry concentration: American manufacturing, 1936-1940', industrial economists have searched for empirical confirmation of the positive relation between profitability and seller concentration. A typical contribution to this literature would regress some version of the Lerner index of market power on some measure of seller concentration (like the Herfindahl-Hirschman index or the n-firm concentration ratio), while controlling for other characteristics of the market (such as advertising intensity and the size of the minimum efficient scale of production).

An example of such a study is the work of Connor and Peterson [1992].[8] They try to explain why there is a difference in price between national brands and comparable private-label products (given that these products are often displayed on adjacent shelves in supermarkets). To test whether these price differences are due to market power of national manufacturers, Connor and Peterson [1992] construct price-cost margins on the basis of these price differences, and regress these on the Herfindahl-Hirschman seller concentration index (adjusted for the own-price elasticity of demand) and two measures of industry advertising intensity, being the advertising-to-sales ratio (*ADBFS*) and the ratio of network television to total advertising expenditures (*TVAD*).[9] In order to control mismeasurement of seller concentration by using only the Herfindahl-Hirschman index, this index is also included as an explanatory variable interacted with indicators for regional markets (*GEOG*), for the share of imports (*IMP*), and for the share of consumers in total demand (*FS*). The model is estimated for two years separately (1979 and 1980), and for the average of these two years.

In Table 6.1 the results of Connor and Peterson [1992] (as they are published) are shown. Discussing these, Connor and Peterson remark (Connor and Peterson [1992, p.164-64, footnote omitted])

> The coefficient of the Herfindahl-Hirschman index of concentration adjusted for the own-price elasticity of demand is positive and significant at the 1% level in all models. As the level of concentration increases or as own-price elasticity of demand decreases (i.e. demand becomes more

[8] An earlier comment on Connor and Peterson [1992] is given by Hinloopen and Martin [1997]. See also Hinloopen and Martin [1995].

[9] Connor and Peterson [1992] argue that (Connor and Peterson [1992, p.158])

> Private-label food manufacturers, on the other hand, operate in markets that have structural configurations that encourage vigorous price competition.

Therefore, the price of private-label products is assumed to equal marginal costs, and the price-cost margin Connor and Peterson [1992] employ is $(p_m - p_c)/p_m$, where p_m is the national brand's price and p_c that of the private label.

Table 6.1   Connor and Peterson's [1992] regression results

	OLS		
	PCM79	PCM80	PCM7980
Constant	-1.5469	7.8919**	7.3439**
	(-0.27)	(3.42)	(3.01)
Concentration $(H/\,|E_d|)$	1.3053**	1.4730**	1.4406*
	(4.31)	(2.74)	(2.42)
Advertising intensity			
ADBFS	2.2882**	1.8870**	2.2074**
	(5.61)	(6.07)	(6.70)
TVAD	0.0993	0.1213*	0.1095*
	(1.41)	(2.14)	(1.73)
Adjustments on Concentration			
Net Imports $(H*IMP)$		-0.7480*	-0.5018
		(-2.25)	(-1.45)
Regional Markets $(H*GEOG)$	-0.6078**	-0.4616**	-0.4891**
	(-2.86)	(-2.71)	(-2.74)
Consumer Products $(H*FS)$	0.5147**	0.3042**	0.3914**
	(3.96)	(2.45)	(3.03)
Growth $(GRO7782)$	0.0672		
	(1.96)		
$\bar{R}^2$	0.51	0.56	0.62

* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

elastic), the price margin between national brands and private label products increases. The significance of $H/\,|E_d|$ is surprising in light of several studies of US manufacturing that found that the concentration-profitability relationship essentially vanished during the inflationary period of the 1970s (Schmalensee [1989, p.975]).

6.3.1 Transcription error, significance levels and alternative results

The results presented by Connor and Peterson [1992] have to be read with caution however. First, it is an established (and good) practise to base *t*-values on White's [1980] heteroscedasticity consistent estimate of the estimated coefficient's standard errors. Heteroscedasticity features especially in cross-sectional studies, such as the one of Connor and Peterson [1992].[10]

---

[10] Indeed, the hypothesis that there is no heteroscedasticity is rejected for every year of the sample.

Table 6.2 Regression Results of OLS and GM[a]

	OLS			GM		
	PCM79	PCM80	PCM7980	PCM79	PCM80	PCM7980
Constant	-1.5469 (-0.28)	7.8919** (5.13)	7.3439** (4.58)	-0.7393 (-0.10)	7.1646** (5.00)	5.9214** (4.49)
Concentration $(HI\|E_d\|)$	-1.3053** (-4.66)	-1.4730** (-4.12)	-1.4406** (-3.28)	-1.6574** (-3.19)	-1.3161** (-3.10)	-1.4910** (-2.58)
Advertising intensity						
ADBFS	2.2882** (4.20)	1.8870** (5.02)	2.2074** (4.60)	2.9751** (4.84)	1.5461** (3.90)	3.0608** (7.07)
TVAD	0.0993* (1.96)	0.1213** (2.55)	0.1095** (2.45)	0.1053* (1.83)	0.1522** (3.50)	0.0937 (1.60)
Adjustments on Concentration						
Net Imports $(H*IMP)$		-0.7480** (-2.71)	-0.5018* (-1.75)		-0.7539 (-1.64)	-0.6293* (-1.98)
Regional Markets $(H*GEOG)$	-0.6078** (-2.62)	-0.4616** (-3.72)	-0.4891** (-2.85)	-0.7999 (-1.53)	-0.4624** (-2.33)	-0.5875* (-2.00)
Consumer Products $(H*FS)$	0.5147** (5.05)	0.3042** (3.94)	0.3914** (4.29)	0.5376* (2.26)	0.3170** (2.56)	0.3653* (2.06)
Growth $(GRO7782)$	0.0672 (1.77)			0.0541 (1.14)		
$\bar{R}^{2b}$	0.51	0.56	0.62	0.78	0.71	0.98

[a] Heteroscedasticity consistent *t*-values in parentheses (White [1980]).
[b] In case of GM estimation, the adjusted goodness of fit indicator is based on weighted data.
* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

Second, and more important, there is a transcription error made by Connor and Peterson. The reported estimated coefficients of the market concentration variable is correct in magnitude, but the true sign is negative, not positive.[11]

In Table 6.2 the true OLS-estimates are reported, together with the concomitant GM estimates. First observe that in most cases the significance levels based on White's [1980] heteroscedasticity consistent standard errors are higher than those reported by Connor and Peterson [1992]. This is especially important for the second variable reflecting advertising intensity, *TVAD*, since it now appears to be (strongly) statistically significant for all years. Indeed, as conjectured by Connor and Peterson [1992], advertising intensity is an important explanatory variable for the difference in price between comparable private label and national brand products. The only variable for which the significance level drops substantially if heteroscedasticity is taken into account is the growth variable,

[11] Connor and Peterson have attributed this to a transcription error when copying the estimation results from the statistical computer package to their text editor (personal communication). See also Connor and Peterson [1997].

GRO7782. But this variable did not significantly explain any part of the price-cost margin anyway (observe that only for the growth variable a two-sided *t*-test is used).

Second, comparing the GM and OLS estimates reveals that in most cases the significance levels of estimated coefficients do not change substantially if outlying observations are accommodated. The most outstanding difference is that the variable *TVAD* looses its significance under GM estimation if the regression is performed on the averaged data (that is, 7980). Also, some of the interactive variables become insignificant when outliers are accommodated (we will come back to these variables in the next subsection). All in all the GM estimates roughly confirm the results generated by OLS, that is, the (correctly reported) results presented by Connor and Peterson [1992] are not heavily influenced by outlying observations.

6.3.2 Nonlinear specification

As mentioned above, Connor and Peterson [1992] have included several variables interacted with the Herfindahl-Hirschman index of seller concentration as explanatory variables. When discussing the effect of seller concentration on profitability, Connor and Peterson [1992] focus on the estimated coefficient of the Herfindahl-Hirschman seller concentration index (adjusted for own-price elasticity). According to the results presented in Table 6.2, it thus appears that the relationship between profitability and concentration is significant and negative.

However, the interpretation of a linear regression with interactive variables is subtle. For instance, the effect of concentration (*H*) on the price-cost margin (*PCM*) in 1979, is given by

$$\frac{\partial PCM79}{\partial H} = \frac{\hat{\beta}_2}{|E_d|} + \hat{\beta}_5 GEOG79 + \hat{\beta}_6 FS79, \tag{6.1}$$

where a hat refers to an estimated value. Indeed, it is incomplete to focus on the coefficient of the linear term alone as characterizing the nature of the concentration-price-cost margin relationship. The concomitant estimated variance is given by

$$var\left[\frac{\partial PCM79}{\partial H}\right] = \frac{var[\hat{\beta}_2]}{E_d^2} + var[\hat{\beta}_5]GEOG79^2 + var[\hat{\beta}_6]FS79^2$$

$$+2cov[\hat{\beta}_2,\hat{\beta}_5]\frac{GEOG79}{|E_d|} + 2cov[\hat{\beta}_2,\hat{\beta}_6]\frac{FS79}{|E_d|} + 2cov[\hat{\beta}_5,\hat{\beta}_6]GEOG79FS79, \tag{6.2}$$

and can be used to validate the significance of (6.1). Observe that both (6.1) and (6.2) differ for every point in the sample.

In Table 6.3 the sign and significance level of (6.1) for the three different regressions, both under OLS and GM estimation, are summarized. Also the magnitude and *t*-value of (6.1) evaluated at the median is given. Although for the majority of observations (6.1) is (significantly) negative for all years, both under GM and OLS estimation, the median value, although negative, does not differ significantly from zero. This means that also in the sample of Connor and Peterson [1992] the concentration-profitability relationship is statistically not present, a result that is not obscured by outlying observations.

Table 6.3 Evaluation of partial derivatives under OLS and GM estimation

	Median	Positive					Negative			
		Total	Significance (2 sided)			Total	Significance (2 sided)		
			1%	5%	10%		1%	5%	10%
						OLS				
$\dfrac{\partial PCM79}{\partial H}$	1.9077 (0.23)	21	8	8	9	24	3	10	12
$\dfrac{\partial PCM80}{\partial H}$	-7.0619 (-1.13)	4	0	0	0	33	12	17	18
$\dfrac{\partial PCM7980}{\partial H}$	-1.8815 (-0.22)	11	0	2	3	23	2	10	12
						GM				
$\dfrac{\partial PCM79}{\partial H}$	-7.5288 (-0.52)	13	0	2	4	32	2	2	6
$\dfrac{\partial PCM80}{\partial H}$	-8.7379 (-0.92)	4	0	0	0	33	1	10	13
$\dfrac{\partial PCM7980}{\partial H}$	-8.9290 (-0.78)	6	0	0	0	28	0	3	11

6.4 Sample-selection procedures

Reviewing the massive literature set in motion by Bain in 1951, Schmalensee [1989] concludes (Schmalensee [1989, p.976])

> The relation, if any, between seller concentration and profitability, is weak statistically, and the estimated concentration effect is usually small. The estimated relation is unstable over time and space and vanishes in many multivariate studies.

Bloch [1994] suggests that Schmalensee's conclusion could be due to erroneous aggregation, since (Bloch [1994, p.71])

> industry classifications do not generally correspond to the boundaries of economically meaningful markets. The inclusion of multiple markets within a single industry classification is clearly discernable at the level of aggregation for data commonly used in cross-industry regressions testing the structuralist hypothesis of a positive relationship between concentration and profitability.

According to Bloch [1994], industry classifications that are highly suspect to include economically different sub-industries are those labelled 'miscellaneous' or 'not elsewhere specified' (nes). He also conjectures that the inclusion of regional industries could blur the empirical relationship between profitability and concentration. According to Bloch [1994] then, controlling for these improper 'industries' would reveal to true (positive and statistically significant) relation between seller concentration and profitability.

6.4.1 Subsample regression results

To test this hypothesis, Bloch [1994] uses data from 1965 and 1980 on gross profit margins (GPMs) and concomitant average Herfindahl-Hirschman indices (H-Hs) for 120 Canadian industries. In particular, Bloch estimates

$$\log(GPM_{1965}/GPM_{1980})_i = \beta_0 + \beta_1 \log(H_{1965}/H_{1980})_i + u_i, \qquad (6.3)$$

where $i = 1, \ldots, 120$, and where u_i is some error term (assumed to be independent and identically distributed with zero mean and bounded variance). The first results reported by Bloch [1994] concern estimates of (6.3) for different subsamples, and are replicated in Table 6.4 (that is, the OLS estimates). Discussing these findings, Bloch remarks (Bloch [1994, p.75, words in parentheses added])

> The estimated coefficient for the concentration ratio in the regression (using the full sample) is positive, but is not statistically greater than zero at the ten percent significance level using a one-tailed t-test. Therefore the structuralist hypothesis that the coefficient of the concentration is greater than zero cannot be accepted in favour of the null hypothesis that the coefficient is less than or equal to zero.

He continues with discussing the results for the different subsamples (Bloch [1994, p.76])

> The estimated coefficient of the concentration ratio in each subsample regression is larger than in the corresponding regression for the full sample. Furthermore, all the estimated coefficients in the subsample regressions are statistically greater than zero at the five percent significance level using a one-tailed t-test.

Bloch concludes (Bloch [1994, p.76])

> Thus, sample selection to exclude industries suspected of containing aggregations of economically meaningful industries alters the estimated relationship between concentration and profitability by an amount sufficient to reverse the conclusion reached on the validity of the structuralist hypothesis.

Bloch's [1994] conclusion may be correct, but they are not as strongly supported by his data as he suggests. In Table 6.4 also the GM estimates for the full sample and the different subsamples are presented. These reveal that for all subsamples *as well as* for the full sample, the relationship between concentration and profitability is statistically significant at the five percent significance level. That is, 'the conclusion reached on the validity of the structural hypothesis' is the same for both the

Table 6.4 Regression results of OLS and GM for full sample and various subsamples of industries (logarithm of (1965/1980) gross profit margins dependent)[a]

Number of Industries	OLS			GM		
	Constant	Logarithm of (1965/1980) H-H indices	$\bar{R}^2$	Constant	Logarithm of (1965/1980) H-H indices	$\bar{R}^{2\,b}$
			Full sample			
120	0.0835[**] (3.27)	0.0836 (1.64)	0.0	0.0431[**] (2.40)	0.0728[*] (1.85)	0.0
		Subsample with only 'miscellaneous' and 'nes' industries omitted				
109	0.0879[**] (3.18)	0.1041[*] (1.77)	0.0	0.0476[**] (2.55)	0.0847[*] (1.87)	0.0
		Subsample with only regional industries omitted				
93	0.0837[**] (2.60)	0.1185[*] (1.97)	0.0	0.0315[*] (1.86)	0.0979[*] (2.35)	0.1
		Subsample with 'miscellaneous', 'nes' and regional industries omitted				
82	0.0925[**] (2.52)	0.1542[*] (2.08)	0.0	0.0385[*] (1.76)	0.1168[*] (2.30)	0.1

[a] Heteroscedasticity consistent t-values in parentheses (White [1980]).

[b] In case of GM estimation, the adjusted goodness of fit indicator is based on weighted data.

[*] Statistically significant at the 5% significance level; [**] Statistically significant at the 1% significance level.

full sample and the different subsamples. Indeed, the GM estimates indicate that Bloch [1994]'s results, especially those for the full sample, are blurred by outlying observations. On the other hand, the GM estimates do confirm that the relationship between profitability and concentration is strongest for those subsamples where aggregation over more than one economically meaningful market is avoided. This is especially true if regional industries are not aggregated.

If anything, the robust estimates indicate that some outlying observations are present in Bloch's [1994] sample. A useful graphical aid for identifying these leverage points and vertical outliers is developed by Rousseeuw and Van Zomeren [1990]. They suggest plotting robust distances (like (5.28) in Chapter 5) against standardized residuals. This plot then allows an exact identification of observations as in Figure 5.1. For Bloch's [1994] full sample this plot is depicted in Figure 6.1. Observations to the right of the vertical line through 2.24 ($= \left[\chi^2_{0.975}(1) \right]^{1/2}$) are leverage points (like points (c) and (d) in Figure 5.1), while points outside the horizontal strip bounded by 3.291 and -3.291 (which corresponds to the 99.99% confidence interval of the standard normal distribution for a two-tailed test) are vertical outliers (as points (b) and (d) in Figure 5.1). Dots in the north-east and south-east rectangles of Figure 6.1 refer to observations which are *both* leverage points and vertical outliers
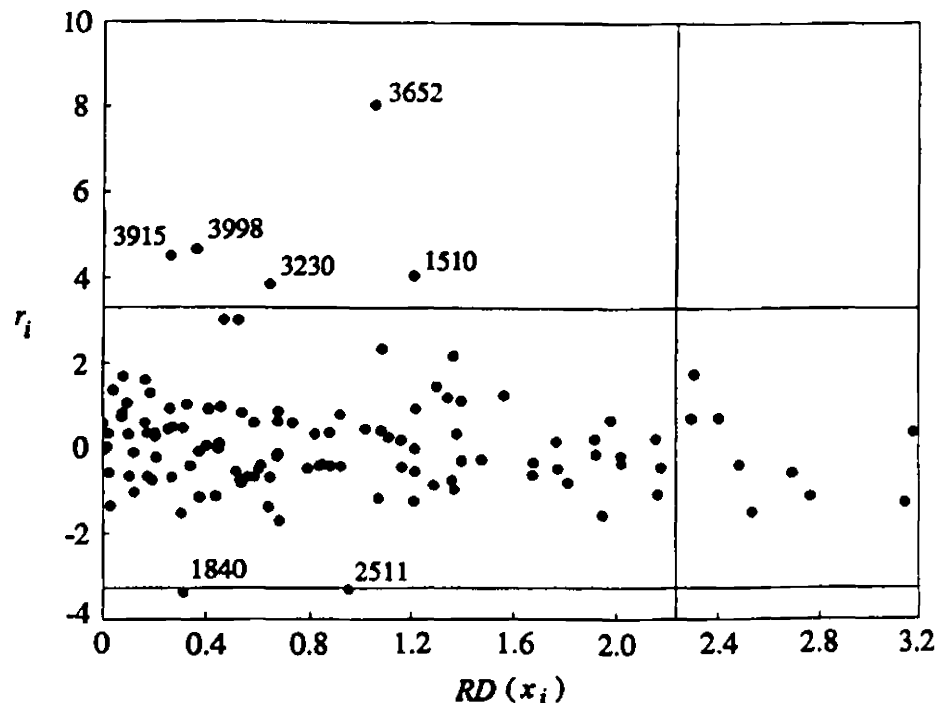
Figure 6.1 Plot of standardized GM-residuals versus robust distances in design space

(like point (*d*) in Figure 5.1). Observe that in Bloch's [1994] full sample there are no such points. Indeed, those leverage points identified through Figure 6.1 coincide with the linear relation of the majority of the data (all leverage points are of type (*c*) in Figure 5.1). On the other hand, vertical outliers abound (that is, observations like (*b*) in Figure 5.1). In Figure 6.1 the SIC-industry code corresponding to these regression outliers are also indicated.[12]

To illustrate the sensitivity of Bloch's [1994] findings to his use of a classic estimator, we have performed some additional OLS-regressions. In particular, twice we have deleted one industry from the full sample (SIC-industry 3652 and 1510), both of which are also present in the three different subsamples. We then re-estimated the relationship between concentration and profitability, the results of which are shown in Table 6.5.[13]

If SIC-industry 3652 is omitted, all estimated magnitudes of the concentration variable and concomitant *t*-values drop substantially. Only the subsamples in which regional industries are omitted reveal a (marginally) statistically significant relation between profitability and concentration. On the other hand, if SIC-industry 1510 is deleted from the sample, *all* estimated relations are statistically significant at the five percent level, and are (much) bigger. Incidently, the result obtained in the latter case for the full sample is statistically stronger than any of the results with SIC-industry 3652 omitted.

---

[12] The 'outlying' industries are 'leaf tobacco processing' (1510), 'cordage and twine industry' (1840), 'shingle mills' (2511), 'motor vehicle manufacturers' (3230), 'manufacturers of lubricating oils and greases' (3652), 'dental laboratories' (3915), and 'fur dressing and dyeing' (3998).

[13] It is quite remarkable that *all* observations which are vertically outlying are neither regional industries nor fall under the headings 'miscellaneous' or 'nes'.

Table 6.5 The sensitivity of OLS; Results of Table 6.4 re-estimated with one observation deleted[a]

Number of Industries	Industry 3652 (SIC) deleted			Industry 1510 (SIC) deleted		
	Constant	Logarithm of (1965/1980) H-H indices	$\bar{R}^2$	Constant	Logarithm of (1965/1980) H-H indices	$\bar{R}^2$
			Full sample			
119	0.0711[**] (3.15)	0.0556 (1.30)	0.0	0.0770[**] (3.04)	0.0971[*] (1.98)	0.02
		Subsample with only 'miscellaneous' and 'nes' industries omitted				
108	0.0735[**] (3.07)	0.0696 (1.43)	0.0	0.0810[**] (3.00)	0.1191[*] (2.09)	0.02
		Subsample with only regional industries omitted				
92	0.0663[**] (2.38)	0.0819[*] (1.72)	0.0	0.0757[**] (2.41)	0.1335[*] (2.30)	0.03
		Subsample with 'miscellaneous', 'nes' and regional industries omitted				
81	0.0711[*] (2.32)	0.1058[*] (1.85)	0.0	0.0840[*] (2.32)	0.1712[*] (2.37)	0.05

[a] Heteroscedasticity consistent t-values in parentheses (White [1980]).
[*] Statistically significant at the 5% significance level; [**] Statistically significant at the 1% significance level.

To summarize this subsection, as observed by Bloch [1994] the statistical relation between profitability and concentration (in his sample) is stronger the less are economically different industries aggregated for the statistical analysis. This is especially true for regional industries. However, Bloch's [1994] sample contains outliers (which typically are not observations from a regional, miscellaneous or 'nes' industries). As revealed by the results presented in Table 6.5 these observations influence (heavily) conclusions based on OLS-regressions. Excluding one industry can either lead to a reduction in all t-values such that the estimates for the different subsamples are at best marginally significant,or can induce all estimates to become statistically significant at (at least) the five percent significance level for a one-sided t-test. Moreover, the estimated magnitude of the relationship between profitability and concentration is also very sensitive to single (possibly outlying) observations. The GM estimator on the other hand is much less influenced by these outlying observations and can be expected to reveal the statistical pattern exhibited by the majority of the data. Applying then this estimator to Bloch's [1994] full sample and subsamples shows that the relation between concentration and profitability is consequent statistically significant. However, also according to the GM estimates this relationship is stronger if regional industries are not grouped together.

6.4.2 Coefficient of variation

Bloch [1994] proposes yet another way to control for the aggregation of economically different industries, motivated by his observation that (Bloch [1994, p.78])

> Aggregation of several homogeneous product markets within an industry classification may also lead to variation in firm size in the industry.

Hence (Bloch [1994, p.79])

> bias in estimates of β may then be limited by selecting samples of industries with low values for the coefficient of variation of firm size.

In Table 6.6 Bloch's [1994] estimates are replicated for samples that differ in variation of firm size, and hence in size (the larger the sample, the higher the coefficient of variation of firm size). Discussing these results, Bloch [1994] concludes (Bloch [1994, p.81, words in parentheses added])

> Support for the structuralist hypothesis of a positive relationship between concentration and profitability is provided from each subsample for which results are shown in Table II (Table 6.6).

Interpreting some of the results presented in Table 6.6 should be done with care however. Especially when the sample is very small, any single observation has large influence on the overall estimate. For example, if SIC-industry 2491 ('fabric glove manufacturers') is excluded from the sample with 15 observations, the t-value of the concentration variable drops to 1.60, which is below the 5% critical value for a one-sided t-test. On the other hand, if from the same sample SIC-industry 1083 ('vegetable oil mills') is deleted, the same t-value rises to 2.59. It is dangerous to rely on (any) statistical procedure if the number of observations is really low.

Focusing then on those samples with, say, at least 30 observations reveals a positive, statistically significant relationship between average seller concentration and profitability (Bloch [1994] reports the results up to the sample with 60 observations together with that comprising 90 industries). But observe that the OLS-estimates for the concentration variable are statistically less strong than the concomitant GM estimates, especially when the sample size is above 100. Again, outliers cloud the relationship between profitability and concentration causing the non-robust estimator to report a too weak relationship. All in all however, Bloch's [1994] second approach to erroneous aggregation seems to be less vulnerable to outlying observations than that discussed in the previous subsection.

Table 6.6 **OLS and GM regression for subsamples of industries with lowest coefficients of variation of firm size (logarithm of (1965/1980) gross profit margins dependent)[a]**

Number of Industries	OLS			GM		
	Constant	Logarithm of (1965/1980) H-H indices	$\bar{R}^2$	Constant	Logarithm of (1965/1980) H-H indices	$\bar{R}^{2\,b}$
4	0.4580* (3.94)	0.9026** (14.04)	0.62	0.4583* (3.80)	0.9024** (14.84)	0.63
5	0.3398 (2.32)	0.8520** (5.52)	0.39	0.3256 (1.75)	0.9949* (3.79)	0.18
7	0.3726* (2.75)	0.9553** (4.04)	0.42	0.3717* (2.59)	0.9887** (3.71)	0.43
10	0.2662* (2.19)	0.7489** (4.13)	0.36	0.0195 (0.16)	0.4134* (2.21)	0.46
15	0.1536 (1.47)	0.4015** (2.08)	0.14	-0.0422 (-0.81)	0.1729 (1.34)	0.15
20	0.1517* (1.88)	0.3483** (2.60)	0.17	0.0498 (0.79)	0.2247* (1.93)	0.17
30	0.0927* (1.82)	0.2822** (2.87)	0.14	0.0152 (0.48)	0.2031* (2.36)	0.23
40	0.0814* (1.78)	0.2496** (2.99)	0.13	0.0371 (1.01)	0.1973** (2.60)	0.14
50	0.0540 (1.29)	0.1679* (2.23)	0.04	0.0116 (0.38)	0.1613** (2.51)	0.11
60	0.0282 (0.77)	0.1260* (1.89)	0.02	0.0003 (0.01)	0.1498** (2.71)	0.13
70	0.0697* (1.71)	0.1962* (2.23)	0.05	0.0253 (1.04)	0.1604** (3.03)	0.13
80	0.0642* (1.85)	0.1724* (2.32)	0.05	0.0253 (1.22)	0.1425** (2.96)	0.12
90	0.0821** (2.51)	0.1711** (2.50)	0.04	0.0297 (1.49)	0.1484** (3.24)	0.13
100	0.0760** (2.55)	0.1295* (2.28)	0.03	0.0286 (1.54)	0.1178** (2.92)	0.10
110	0.0760** (2.78)	0.1004 (1.48)	0.01	0.0290 (1.57)	0.0950* (2.32)	0.06
120	0.0835** (3.27)	0.0836 (1.64)	0.01	0.0431** (2.40)	0.0728* (1.85)	0.04

[a] Heteroscedasticity consistent *t*-values in parentheses (White [1980]).
[b] In case of GM estimation, the adjusted goodness of fit indicator is based on weighted data.
* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

6.5 Cyclical variation in the profit-concentration relationship

Schmalensee's [1989] stylized fact, that the concentration-profitability relation is at best weak statistically, is also challenged by Mueller and Sial [1993]. They conjecture that many empirical investigations to the concentration-profitability relation are based on data taken from a time-period in which cyclical factors seriously distort the alleged relationship. Hence, if these cyclical factors are controlled for the concentration-profitability relation should be retrieved empirically.

In order to test their hypothesis, Mueller and Sial [1993] use US data consisting of annual observations on 17 SIC-industries for 33 years. The observed variables include profits (*P*), four firm concentration ratio (*C4*), imports (*IM*), change in demand (*G*), unionization rate (*UN*), capacity utilization rate (*CU*), unemployment rate (*U*), and producer inflation rate (*PI*).

6.5.1 Reported results

In order to illustrate the distorting effect of cyclical factors on the concentration-profitability relationship, Mueller and Sial [1993] first estimate the relation between profit rates and concentration for each year of their sample separately.[14] When discussing the results of these annual regressions Mueller and Sial [1993] distinguish three periods. For the first period (Mueller and Sial [1993, p.285, words in parentheses added])

> The *t*-value of the coefficient on *C4* is statistically significant at the 5% level in at least one model (the most elementary) in all but five years during 1947-1972.

For the second period, Mueller and Sial [1993, p.285] conclude

> The significant positive relationship between profits and concentration disappears in all but two years during 1973-1981.

As for the last period, Mueller and Sial [1993, p.285, words in parentheses added] remark

> During 1982-1990, a significant positive relationship between profit and concentration again manifested itself in all but three years (1987, 1988 and 1989). This was a period of moderate inflation accompanied by quite high unemployment rates. While the results for 1989 may reflect the relatively high inflation rate for that year, we have no explanation for the absence of a significant positive relationship in 1987 and 1988.

Mueller and Sial [1993] proceed with estimating the pooled cross-sections. Their OLS and GLS results are replicated in Table 6.7. On the basis of these findings Mueller and Sial [1993, p.287, footnote omitted] observe

---

[14] In addition, Mueller and Sial [1993] also report, for every year, estimates of the relationship between profits and (i) concentration, change in demand, and unionization rate, and (ii) concentration, change in demand, unionization rate and imports. Their main focus however is on the estimated coefficient of *C4*. Indeed, in all annual regressions it is only the estimate of this coefficient Mueller and Sial [1993] report.

Table 6.7 OLS and GLS estimates of pooled regressions (Mueller and Sial [1993, Table II])[a]

	OLS (1A)	GLS (1B)	OLS (2A)	GLS (2B)	OLS (3A)	GLS (3B)
Constant	11.80	16.72	-11.93	-14.76	-6.08	-11.94
C4	0.20**	0.08**	0.22**	0.06*	0.15**	0.03
	(10.28)	(2.45)	(11.98)	(1.70)	(4.84)	(0.76)
G	0.17**	0.17**	0.14**	0.07**	0.12**	0.06**
	(10.76)	(10.44)	(7.23)	(3.19)	(6.59)	(2.90)
IM	-0.38**	-0.30**	-0.42**	-0.37**	-0.39**	-0.33**
	(9.42)	(5.68)	(10.63)	(6.94)	(10.19)	(6.16)
UN	-6.36**	-5.02*	-5.55**	-5.36*	-6.71**	-6.01**
	(4.28)	(2.14)	(3.95)	(2.15)	(4.91)	(2.41)
CU			0.21**	0.35**	0.07	0.25**
			(4.51)	(6.25)	(1.35)	(3.97)
U			0.76**	0.68**	0.60**	0.57**
			(3.63)	(3.54)	(2.95)	(3.05)
DU			-1.14**	-0.91**	-1.10**	-0.87**
			(4.33)	(3.84)	(4.40)	(3.80)
PI			0.36**	0.33**	1.56**	1.42**
			(7.43)	(7.79)	(7.60)	(7.73)
C4×PI					-0.02**	-0.02**
					(6.08)	(6.16)
C4×CU					0.002**	0.002**
					(5.61)	(3.80)
R^2	0.36	0.32	0.46	0.35	0.51	0.42

[a] *t*-values in parentheses.

* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level

The coefficients on all variables in all GLS models have the expected signs and, except for the C4 coefficient in 3B, are significant at the 5% level or higher...inclusion of cyclical variables, CU, U, DU, and PI has a significant independent effect on profitability...including C4*PI and C4*CU decreased the coefficients for the independent variables C4 and CU...The expected negative sign on the coefficient for C4*PI indicates that as the inflation rate increases, the profits of less concentrated industries rise relative to those of concentrated industries. In contrast, the expected positive coefficients for C4*CU show that as the rate of capacity utilization rises the profits of high concentration industries rise relative to low concentration industries.

Finally, Mueller and Sial [1993] explore in detail the relationship between concentration and profitability for different levels of inflation. In particular, the predicted profit levels (on the basis of the pooled cross-section GLS regression) are calculated for different levels of concentration and infla-

Table 6.8 Predicted profit rates at various concentration levels and inflation rates (Mueller and Sial [1993, Table III])

Four-firm Concentration	Inflation rate						
	0	2.4	5	10	15	20	22
$C4 = 80$	24.0	23.6	23.1	22.2	21.3	20.4	20.1
$C4 = 55$	19.2	19.9	20.8	22.4	24.0	25.6	26.2
$C4 = 30$	14.3	16.3	18.4	22.5	26.6	30.7	32.3
$C4\mid_{=80} - C4\mid_{=30}$	9.7	7.3	4.7	-0.3	-5.3	-10.3	-12.2

tion (all other explanatory variables are set at their mean value). In Table 6.8 Mueller and Sial's[1993] predicted profitability levels are depicted. According to Mueller and Sial [1993, p.288] these results indicate that

> As expected, the difference in the profit rate of a $C4=80$ industry and a $C4=30$ industry decreases as the inflation rate increases...In sum, these findings suggest why in annual cross sectional equations there was not a significant positive relationship between profits and $C4$ in most years during 1974-1981.

Although the analysis of Mueller and Sial [1993] contains interesting observations indeed, their results should be read with care. It will appear that several findings of Mueller and Sial [1993] are dictated by a few outlying observations.

6.5.2 Corrected OLS regression results

On the basis of the data supplied by Willard Mueller, we have been able to replicate exactly the (117!) annual OLS regressions, as well as the OLS regressions for the first two equations of Table 6.7. The GLS estimates however could not be replicated, since Mueller and Sial [1993] do not indicate which specific GLS estimator they use.[15] More importantly however is that the OLS regression result for the third equation of Table 6.7 differ markedly from those dictated by the data.[16]

In Table 6.9 the 'true' OLS regression results are presented (ignore for the moment the GM estimates). The main difference between these and those reported by Mueller and Sial [1993] is that

---

[15] Mueller and Sial [1993] refer to some statistical computer package which (apparently) provides some default GLS estimator.

[16] The probability of there being some errors in the data is practically zero, since the only difference between the second and third equation in Table 6.7 is the inclusion of two interactive terms. Hence, equation three is based on exactly the same data as equation two.

Table 6.9 True' OLS and GM estimates of pooled regressions[a]

	OLS (1A)	GM (1B)	OLS (2A)	GM (2B)	OLS (3A)	GM (3B)
Constant	11.8027**	12.7653**	-11.9325*	-0.5548	-11.5923	14.7213
	(10.24)	(11.28)	(-2.09)	(-0.14)	(-0.80)	(1.53)
C4	0.2045**	0.1937**	0.2238**	0.1989*	0.2254	-0.0842
	(9.31)	(8.89)	(10.17)	(9.38)	(0.96)	(-0.45)
G	0.1741**	0.1393**	0.1379**	0.1401**	0.1315**	0.1384**
	(8.72)	(9.08)	(6.74)	(7.83)	(6.67)	(7.95)
IM	-0.3836**	-0.3749**	-0.4179**	-0.4545**	-0.4033**	-0.4248**
	(-6.61)	(-6.47)	(-7.56)	(-7.12)	(-7.74)	(-6.98)
UN	-6.3642**	-5.02*	-5.5523**	-4.3931*	-5.7759**	-4.9398**
	(-4.69)	(2.14)	(-4.15)	(-3.51)	(-4.54)	(-4.17)
CU			0.2076**	0.0842*	0.1613	-0.1174
			(3.32)	(1.98)	(0.97)	(-1.08)
U			0.7638**	0.7186**	0.7088**	0.6142**
			(4.13)	(4.12)	(3.90)	(3.58)
DU			-1.1360**	-1.2901**	-1.1224**	-1.1804**
			(-4.65)	(-5.75)	(-4.68)	(-5.30)
PI			0.3613**	0.3860**	1.1941**	0.9530**
			(6.47)	(7.72)	(6.33)	(4.77)
C4×PI					-0.0168**	-0.0112**
					(-4.60)	(-2.62)
C4×CU					0.0009	0.0039**
					(0.34)	(1.77)
R^2						

[a] Heteroscedasticity consistent t-values in parentheses (White [1980]) under OLS-estimates.
* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

both the estimated coefficient for C4 and C4×CU are *not* statistically significant.[17] Hence, the estimate does *not* confirm that 'as the rate of capacity utilization rises the profits of high concentration industries rise relative to low concentration industries'. Whether or not the GLS estimate of the third equation, as presented by Mueller and Sial [1993], is correct, remains to be checked.

[17] Observe that Mueller and Sial [1993] report conventional t-statistics while those given in Table 6.9 are consistent under heteroscedasticity. However, also the conventional t-statistics indicate that the estimated coefficients of both C4 and C4×CU statistically do not differ from zero significantly.

Table 6.10 Predicted profit rates at various concentration levels and inflation rates based only on significant explanatory variables of the GLS estimate

Four-firm Concentration	Inflation rate						
	0	2.4	5	10	15	20	22
$C4$ = 80	21.6	17.8	13.6	5.6	-2.4	-10.4	-13.6
$C4$ = 55	17.5	14.9	12.0	6.5	1.7	-4.5	-6.7
$C4$ = 30	13.4	11.9	10.4	7.4	4.4	1.4	0.2
$C4\|_{=80} - C4\|_{=30}$	8.2	5.9	3.2	-1.8	-6.8	-11.8	-13.8

6.5.3 Alternative results

As mentioned above, Mueller and Sial [1993] predict profit rates on the basis of the third pooled cross-section GLS regression as presented in Table 6.7 (equation 3B), see Table 6.8. Apart from the fact that the GLS estimate might not be accurate, observe that the predicted profit rates are based on *all* explanatory variables included in the GLS regression, including those which are not statistically significant. But this can have important consequences. If some observable is to be explained by a linear regression, and an explanatory variable with positive mean, which has nothing to do with the dependent observable, is added to this regression, the concomitant estimated coefficient could easily be different from zero in magnitude, but not in statistical significance. Including then this added explanatory variable when predicting the dependent observable (on the basis of the regression results) leads to biased predictions. Indeed, it is this bias which Mueller and Sial [1993] introduce when predicting profitability on the basis of all regressors.

In Table 6.10 the predicted profit rates are presented based on Mueller and Sial's [1993] GLS estimate, but only including those explanatory variables which are statistically significant. Comparing the results of Table 6.10 with those of Mueller and Sial [1993] (see Table 6.8) shows that indeed the predicted values of Mueller and Sial [1993] are seriously biased. Not only are now all predicted values (much) lower, but for all concentration levels profits decline when inflation goes up, all else equal. On the other hand, the main conclusion of Mueller and Sial [1993], that 'the difference in the profit rate of a $C4$ = 80 industry and a $C4$ = 30 industry decreases as the inflation rate increases' is still true. However, the validity of this exercise is still to be questioned, since, again, there is reason to doubt the correctness of the GLS estimate as presented by Mueller and Sial [1993], on which both the results of Tables 6.8 and 6.10 are based.[18]

---

[18] The predicted profit rates based on the 'true' OLS regression results, including only those explanatory variables which are statistically significant, are also always decreasing in the inflation rate. Moreover, the difference between profit rates when concentration is high ($C4$ = 80) and low ($C4$ = 30) is always negative when the inflation rate is positive. If anything, these results do not 'suggest why in annual cross sectional equations there was not a significant positive relationship between profits and $C4$ in most years during 1974-1981'.
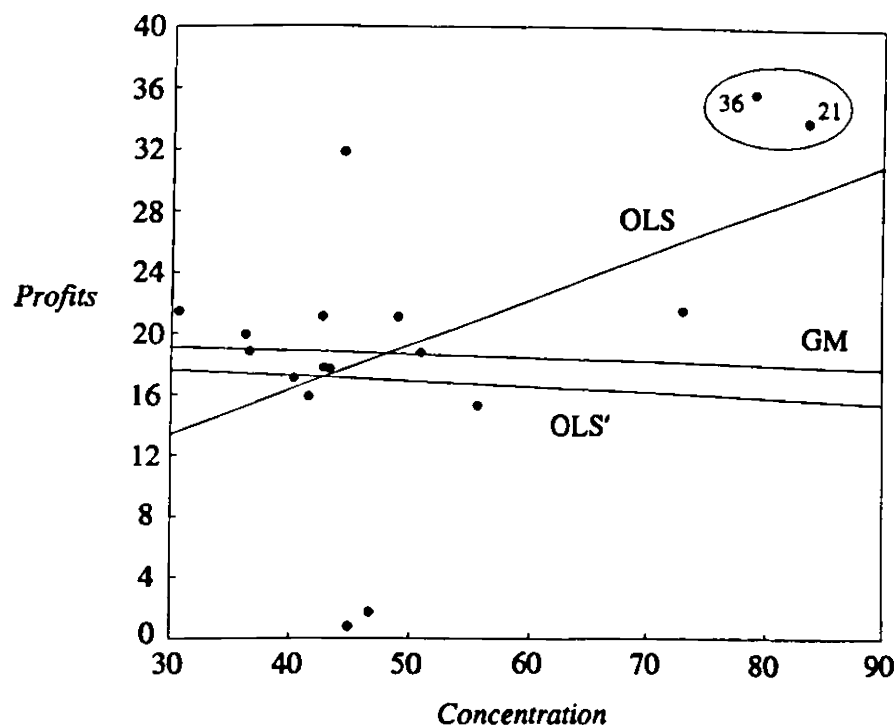
Figure 6.2 Concentration versus profits; 1984

6.5.4 GM regression results

To examine the effects, if any, of outlying observations on the results presented by Mueller and Sial [1993], consider first the annual regressions. Observe that each of these regressions are based on only 17 observations. As mentioned in Section 6.4.2, any statistical analysis based on so few data points should be considered with care. The analysis of Mueller and Sial [1993] is no exception.

A typical result is the annual regression for 1984, which reads[19]

$$P_i = 4.8131 + 0.2959\,C4_i, \qquad\qquad (6.4)$$
$$ (0.89) \quad (3.03)$$

where the coefficients are estimated with OLS (*t*-values based on White's [1980] heteroscedasticity consistent estimate of the coefficients' standard errors are in parentheses). The estimated coefficient of $C4$ appears to be highly significant, a result which emerges frequently for the period 1982-1990 (as stressed by Mueller and Sial [1993]). Still, the regression is based only on 17 observations, raising the suspicion that OLS results may be influenced by (possible) outlying observations.

In Figure 6.2 the 1984 four-firm concentration ratios for the 17 SIC-industries are plotted against the corresponding profit rates. Also the estimated OLS-line is drawn (ignore for the moment

---

[19] The estimated coefficient of $C4$ for the other two regressions (see Section 6.5.1) in 1984 are respectively 0.3987 and 0.3575, with concomitant (heteroscedasticity consistent) *t*-values equal to 4.64 and 7.05.
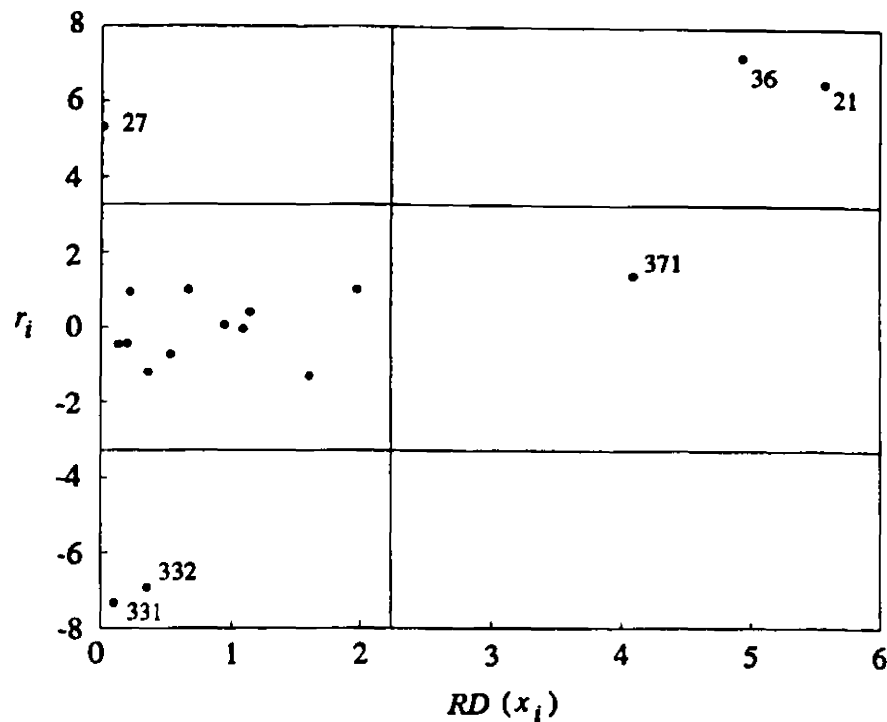
Figure 6.3 Standardized GM-residuals versus robust distances in design space; 1984

the other two lines). Although OLS results show a positive and statistically significant estimate of the slope-parameter, a subjective judgement, based on the spread of observations, could be that for the majority of the data, the OLS-estimate is off the mark. Indeed, if the two circled observations are ignored (which in this case amounts to the rejection of almost 12% of the data), the statistical relation could well be negative, or nonexistent.[20]

To examine the OLS estimate further, consider the relevant GM estimate, which is given by

$$P_i = 19.7622 - 0.0227\,C4_i,$$
$$(4.54)\quad(-0.22)$$

(6.5)

and which is also drawn in Figure 6.2. This result is strikingly different. Not only is the sign of the concentration variable reversed, but the estimate also loses its statistical significance. This suggests that some outlying observations corrupt the data and the OLS estimate.

Further examination of the data reveals that indeed a few outliers distract OLS. In Figure 6.3 the robust (MVE) distances of the explanatory variables are plotted against the standardized GM-residuals (compare Figure 6.1). There appear to be five vertical outliers, two of which are also leverage points (SIC-industry 21 and 36).[21] It is the latter two observations which attract the OLS-line (and

---

[20] The corresponding industries are 'tobacco products' (21), and 'electronic & other electric equipment' (36).

[21] The other industries are 'printing and publishing' (27), 'primary iron and steel' (331-332; these two industries appear under the same heading, see Mueller and Sial [1993, p.282]), and 'motor vehicles and equipment' (371).

which are circled in Figure 6.2). Indeed, if these are removed from the data set, OLS yields

$$P_i = 17.9582 - 0.0128C4_i,$$
$$(3.58) \quad (-0.12)$$

(6.6)

which is also depicted in Figure 6.2. Rejecting the two outlying observations not only reverses the sign of the estimated coefficient of the concentration variable, it now also appears statistically not to differ significantly from zero. In fact, the OLS and GM estimate almost coincide.

Mueller and Sial's [1993] discussion of the annual regressions is correct given their estimates. But these estimates could well be highly influenced by a few outlying observations. Considering the OLS estimate for 1984 when two outliers are rejected challenges the observation that 'during 1982-1990, a significant positive relationship between profit and concentration again manifested itself in all but three years'.[22]

But outlying observations also trouble the OLS regression results of Mueller and Sial [1993] if we consider the whole sample, although to a much lesser extent. Comparing the GM and OLS results of the first pooled cross-section (see Table 6.7) reveals that both techniques find all coefficients to be statistically significant. The main difference is in the estimated size of the 'change in demand' coefficient (G). The magnitude of the GM estimate is some 80% of the OLS estimate. The differences for the second equation are more pronounced. Not all coefficients are found to be of equal statistical significance by the two estimation techniques. And the estimated magnitude of the CU-coefficient under GM estimation is just 40% of the corresponding OLS estimate. The most marked differences however are in the third equation. The estimated signs for the concentration and capital utilization coefficients are now opposite. Moreover, the second interactive variable, $C4 \times PI$, is found to be (marginally) significant when outlying observations are accommodated, whereas it appears to be statistically insignificant when OLS is used.

Again, the differences between the OLS and GM estimates can be attributed to outliers. The familiar residual-distance plot is presented in Figure 6.4 for the second equation of Table 6.7 (the plot for the first and third equation appears to be quite similar).[23] Several observations are identified as being vertically outlying, while seven of these are also leverage points. In Figure 6.4 all these outlying observations are labelled with their year and SIC-code (note that there are numerous observations which are leverage points but not vertical outliers). It is interesting to observe that SIC-industry 331, which is a vertical outlier and leverage point in 1982, 1983 and 1986, is also 'double' outlying in equations 1 and 3.

In Table 6.11 the OLS-estimates are presented when SIC-industry 331 is deleted from the sample for the years 1982, 1983 and 1986 (this implies a reduction in the number of observations of less than 0.6%). As expected the estimate of the first equation remains practically the same. The second equation shows more signs of influence. In particular, the estimated size of the capital utilization variable is almost halved. Not surprisingly, given the GM estimate, the third equation suffers most from outliers. The sign of the estimated coefficient for the second interactive term,

---

[22] In fact, the estimated relationship between profitability and concentration for these years was statistically strongest in 1984.

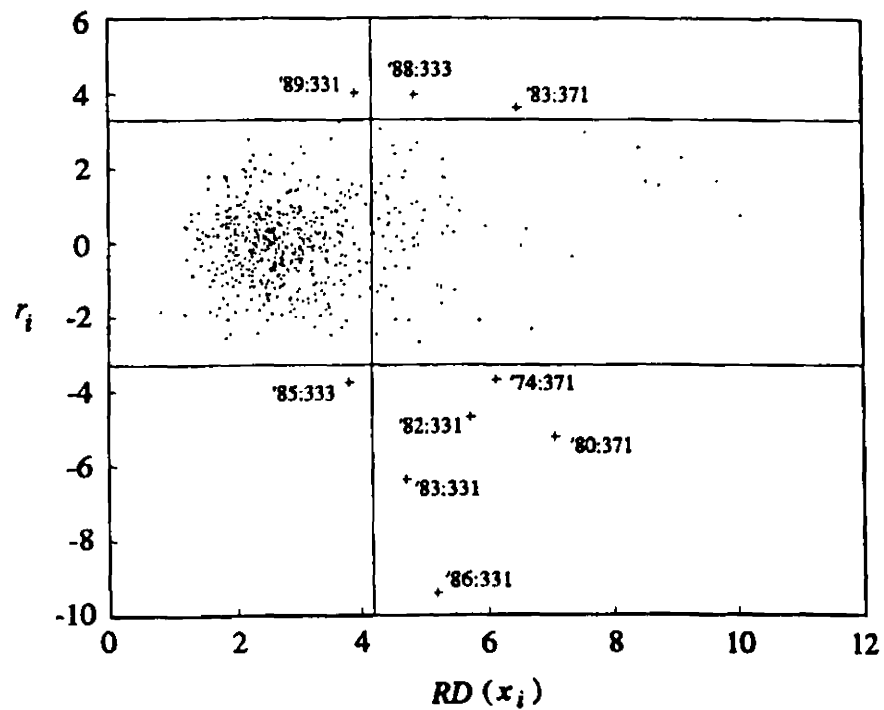[23] Industry 333 refers to 'primary nonferrous metals'.

Figure 6.4 Standardized GM-residuals versus robust distances in design space; whole sample

$C4 \times CU$ is reversed, and the statistical significance of the cyclical variable U drops substantially. More important however is the enormous increase in significance of the concentration variable. The rejection of less than 0.6% of the data induces the (heteroscedasticity consistent) t-value to increase from 0.96 to 4.89. Indeed, also in relatively big data sets a few outliers can seriously obscure classic estimates of coefficient size and statistical significance.

Table 6.11 OLS estimates of pooled regressions; SIC 331 deleted in 1982, 1983 and 1986[a]

	OLS (1A)	OLS (2A)	OLS (3A)
Constant	12.2720**	-4.7547	-9.19533*
	(11.23)	(-1.14)	(-1.84)
C4	0.1923**	0.2063**	0.2995**
	(9.20)	(10.38)	(4.89)
G	0.1534**	0.1371**	0.1320**
	(9.61)	(7.50)	(7.68)
IM	-0.3432**	-0.3928**	-0.3801**
	(-6.78)	(-7.72)	(-8.15)
UN	-5.4087**	-4.5500**	-4.7864**
	(-4.36)	(-3.75)	(-4.22)
CU		0.1290**	0.1307**
		(2.87)	(3.05)
U		0.7505**	0.8442*
		(4.13)	(1.75)
DU		-1.2720**	-1.2585**
		(-5.68)	(-5.73)
PI		0.3295**	1.1395**
		(6.48)	(6.01)
C4×PI			-0.0164**
			(-4.22)
C4×CU			-0.0027
			(-0.28)
R^2	0.34	0.44	0.46

[a] Heteroscedasticity consistent *t*-values in parentheses (White [1980]).
* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

6.6 Firm entry and firm growth

An example of empirical research in Industrial Organization not concerned with the traditional concentration-profitability relationship is the study of Mata [1993]. Using data from 1982, consisting of a sample of 65 Portuguese industries, Mata [1993] examines whether 'firm entry' and 'firm growth' react qualitatively and quantitatively in the same way to market characteristics.

An empirically testable model is constructed in which firm entry (ENTRY) and firm growth (EXPANSION), measured respectively as the entering or expanding *number of firms* or *number of*

Table 6.12 OLS and GM regression results[a]

	Number of Firms				Employment			
	Entry (1)		Expansion (2)		Entry (3)		Expansion (4)	
	OLS	GM	OLS	GM	OLS	GM	OLS	GM
CONST	16.64	-1.75**	10.65*	-0.86	16.25	-2.69**	16.18	-1.85**
	(1.56)	(-2.55)	(1.66)	(-1.14)	(1.38)	(-2.32)	(1.44)	(-3.60)
PROFIT	6.82*	0.51	5.78*	-0.28	2.35	0.92	1.74	-0.03
	(1.98)	(1.07)	(1.98)	(-0.65)	(0.86)	(0.65)	(0.59)	(-0.04)
GR	-0.50*	-0.02	-0.42*	-0.02	-0.28	-0.04	-0.33	-0.05
	(-1.79)	(-0.28)	(-1.81)	(-0.34)	(-1.10)	(-0.22)	(-1.17)	(-1.01)
LSIZE	1.82**	0.28**	1.72**	0.33**	1.60**	0.51**	1.66**	0.26**
	(4.10)	(2.31)	(5.79)	(4.72)	(3.75)	(2.82)	(3.92)	(5.79)
LMES	-1.83**	-0.16	-1.57**	-0.23**	-1.05*	-0.25	-0.94*	-0.16**
	(-3.31)	(-1.48)	(-4.21)	(-3.86)	(-2.00)	(-1.56)	(-1.79)	(-2.78)
LKL	-0.84*	-0.06	-0.65*	-0.06*	-0.80*	-0.00	-0.91*	-0.02
	(-1.84)	(-1.59)	(-2.08)	(-1.73)	(-1.65)	(-0.03)	(-1.84)	(-0.58)
EQUIP	-4.68*	0.28*	-3.27*	0.08	-5.40*	0.17	-5.32*	0.33*
	(-1.68)	(1.68)	(-1.96)	(0.46)	(-1.76)	(0.45)	(-1.82)·	(2.28)
NEW	-2.17	0.23	-1.98*	-0.05	-2.03	0.04	-2.60	-0.01
	(-1.53)	(0.84)	(-1.87)	(-0.21)	(-1.32)	(0.07)	(-1.62)	(-0.04)
ADV	-1.26*	-0.35**	-0.86	0.02	-1.17**	-0.45*	-0.64	0.23
	(-1.95)	(-3.00)	(-1.61)	(0.29)	(-2.46)	(-2.22)	(-1.21)	(0.84)
PAT	-4.23	-0.62	-4.33*	-0.74*	-1.15	-1.76*	-1.79	-0.53
	(-1.35)	(-0.91)	(-1.69)	(-1.71)	(-0.36)	(-1.94)	(-0.78)	(-0.97)
$\bar{R}^{2}$[b]	0.46	0.56	0.56	0.70	0.42	0.62	0.43	0.70

[a] Heteroscedasticity consistent *t*-values in parentheses (White [1980]).
[b] In case of GM estimation, the adjusted goodness of fit indicator is based on weighted data.
* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

employees in those firms from 1982 to 1986, is related to a measure of expected profitability (PROFIT), to an entry barrier (the minimum efficient scale size (MES), which is interacted with industry size (SIZE) and the capital/labour ratio (KL)), to two indicators of sunk costs (the logarithm of the expected life of machines (EQUIP) and an inverse measure of the ease of reselling capital goods (NEW)), and to two measures of product differentiation (the average (1979-82) percentage of the

patent and trademark expenditures in total production (PAT) and a dummy (ADV) being 1 if the industry is consumer orientated in Portugal and advertising intensive in Spain).

6.6.1 OLS and GM regression results

The regression results of Mata [1993], as well as the corresponding GM estimates, are presented in Table 6.12.[24] As to the equality of the determinants for firm entry and firm growth, Mata [1993, p.575] remarks (on the basis of the OLS regression results)

> Comparison of results in columns (1)-(2) and (3)-(4) shows that entry and growth respond in the same way to the same stimuli, which suggests that they may be close substitutes, in the sense that firm entry and firm growth may be alternative ways to absorb industry excess profits, and that the choice between these alternatives may not depend on the height of entry/mobility barriers.

Mata [1993] also tests formally for the equivalence of the two regressions. And indeed, at the 5% significance level joint equality of coefficients can not be rejected.

Turning now to the GM estimates a number of observations are to be made. In terms of significance, the variables LSIZE, NEW, ADV and PAT are roughly the same under OLS and GM estimation. This means in particular that the significant explanatory power of the size of an industry for firm growth and firm expansion is a robust finding. On the other hand, the estimated magnitude of the coefficients are strikingly smaller when outlying observations are accommodated (except for the PAT-variable in the entry equation measured in numbers of employees). As will be shown below, in this respect the OLS regression results are heavily influenced by a few outlying observations.

Second, under GM estimation the variables PROFIT and GR lose their significance in the equations based on the number of firms. For the GR variable this means that an unexpected sign is found to be insignificant. Further, the logarithm of the capital-labour ratio (LKL) remains only (marginally) significant in the expansion equation based on the number of firms. Indeed, the estimated coefficient of this ratio (based on OLS), with corresponding t-value, appears to be heavily influenced by a few outliers.

Third, a striking difference between the classic and robust estimate is the sign reversal of the estimated coefficient for the first sunk cost variable, EQUIP. According to the GM estimate, the longer the expected life of machines and other equipment will be, the higher the number of expanding and growing firms (this effect however is only found to be truly statistically significant in the growth equation based on the number of employees in expanding firms). Again, the OLS estimates for the variable EQUIP appear to be determined to a great extent by only a few observations (see below).

Finally, according to the GM estimates the (logarithm) of the minimum efficient scale size (LMES) does not constitute a barrier to entry, while it appears to be a major barrier to growth.

---

[24] Observe that the interpretation of the coefficients of the logarithmic variables should be done with care. The calculated t-values are correct, but the estimated size of the effect of the logarithmic designs is given by the appropriate partial derivatives, which differ for every point in the sample (see also Subsection 6.3.2).
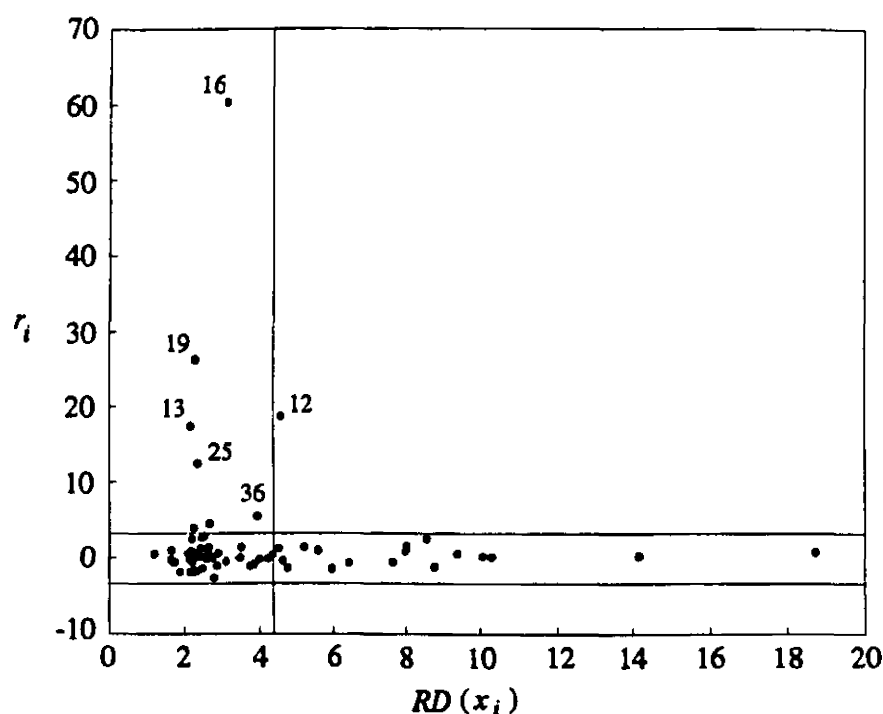
Figure 6.5 Standardized GM-residuals versus robust distances in design space; eq. (1), Table 6.12

6.6.2 Alternative OLS regression results

The differences between Mata's [1993] results and those obtained when using the robust estimator can be attributed to outlying observations. In Figure 6.5 the standardized GM-residuals of the first equation of Table 6.12 are plotted against the robust distances of the design matrix (compare Figure 6.1). Mata's [1993] sample contains some major vertical outliers as well as numerous (good) leverage points.[25] The three most severe vertical outliers (observations 13, 16 and 19) appear to be vertically outlying in all four estimated equations. Observe that these three observations together constitute less than 5% of the whole sample.

In Table 6.13 OLS regression results are shown for the sample in which observations 13, 16 and 19 are deleted. The estimated coefficients and concomitant *t*-values are substantially different from those reported by Mata [1993]. The two sunk costs variables (EQUIP and NEW) as well as the capital-labour ratio (LKL) lose their significance, and the minimum efficient scale (LMES) appears only to be a barrier to entry and expansion if either of these is measured as the number of firms (and not as the number of employees in those firms). Moreover, the magnitude of most estimated coefficients drops substantially (except for PROFIT, GR, and, to some extent, PAT), in some cases even dramatically.

---

[25] The vertical outliers are 'spinning and weaving of man-made fibres' (12), 'manufacture of knit fabrics' (13), 'ready made clothing' (16), 'footwear' (19), 'wooden furniture' (25), and 'tires' (36).

Table 6.13 OLS regression results; three vertical outliers deleted[a]

	Number of Firms		Employment	
	Entry (1)	Expansion (2)	Entry (3)	Expansion (4)
CONST	0.6920	-1.5471	-3.2421	-4.7231*
	(0.23)	(-0.56)	(-1.37)	(-1.83)
PROFIT	6.9931*	5.8627*	2.4160	1.7951
	(2.23)	(2.28)	(1.36)	(1.14)
GR	-0.4480*	-0.3808*	-0.2112	-0.2605*
	(-1.85)	(-2.17)	(-1.22)	(-1.94)
LSIZE	1.2219**	1.2516**	0.8598**	0.8550**
	(3.21)	(4.34)	(3.88)	(4.44)
LMES	-1.1715**	-1.0926**	-0.2730	-0.1219
	(-2.61)	(-3.27)	(-1.30)	(-0.74)
LKL	-0.1154	-0.0810	0.0651	0.0349
	(-1.06)	(-0.80)	(0.65)	(0.36)
EQUIP	-0.4954	-0.0658	-0.2476	0.1982
	(-0.64)	(-0.09)	(-0.60)	(0.43)
NEW	-1.3722	-1.2618	-0.8026	-1.2419*
	(-1.09)	(-1.41)	(-1.06)	(-1.71)
ADV	-1.0902	-0.6867	-0.9236**	-0.3598*
	(-1.58)	(-1.15)	(-2.78)	(-1.70)
PAT	-4.9059*	-4.7952**	-1.8342	-2.4993*
	(-2.23)	(-2.51)	(-1.05)	(-2.02)
$\bar{R}^{2\,b}$	0.38	0.49	0.38	0.43

[a] Heteroscedasticity consistent *t*-values in parentheses (White [1980]).
[b] In case of GM estimation, the adjusted goodness of fit indicator is based on weighted data.
* Statistically significant at the 5% significance level; ** Statistically significant at the 1% significance level.

On the basis of Table 6.13 it is more difficult to maintain that firm entry and firm growth 'respond in the same way to the same stimuli'. That is, Mata's [1993] confirmation of the equivalence hypothesis may well be based upon only a few outlying observations.

To conclude this section, the results of Mata [1993], although very suggestive, are also sensitive to outlying observations. If less than five percent of the sample is rejected quite different empirical findings emerge. Not surprisingly, the GM-estimates are rather different from those reported by Mata [1993]. The main result appears to be that both the GM estimates as well as the augmented OLS results appear to be inconsistent with the conclusion that firm entry and firm growth respond qualitatively and quantitatively in the same way to the same market characteristics.

6.7 Conclusions

Two decades ago, Almarin Phillips wrote 'A critique of empirical studies of relations between market structure and profitability'. He concludes his critique with three necessities (Phillips [1976, p.248])

> What do we know from the empirical studies of relations between structure and profitability? Very little, it appears...Better theory, better data, and above all, better econometrics are needed before policy can be based on anything other than in-depth institutional studies of particular markets.

The present chapter documents an attempt to use 'better econometrics' in the sphere of Industrial Organization. In particular, the robust GM estimator developed in Chapter 5 is used to re-examine alleged statistical relationships, as they have been published in four studies over the last five year. This exercise allows two conclusions.

The first observation concerns care in reporting results. Two out of the four studies considered here reported incorrect results. Careful scientific practise should effectively preclude this type of error.

The second observation is that in all four studies considered, the use of OLS was shown to yield estimates which are (strongly) dictated by only a few outlying observations. In some cases, the rejection of only a small percentage of the data was shown to have a substantial effect in the OLS regression results. Conclusions based on these classic regressions results should therefore be drawn with care. On the other hand, the robust GM estimator yields much more reliable estimates. Indeed, it is these regression results which reveal the statistical relation contained in the majority of the data.

The limited scope of the present chapter should be emphasized. The whole analysis is based on only four studies. Future research should reveal more firmly whether or not the use of classic estimators has led in general to statistical digressions.

References

Anderson, S.P., 1987, "Spatial competition and price leadership", *International Journal of Industrial Organization*, 5, 369-95.

Anderson, S.P., 1988, "Equilibrium existence in the linear model of spatial competition", *Economica*, 55, 479-91.

Anderson, S.P., de Palma, A. and Thisse, J.-F., 1992, *Discrete choice theory of product differentiation*, Cambridge: MIT Press.

Anderson, S.P. and Engers, M., 1994, "Spatial competition with price-taking firms, *Economica*, 61, 125-36.

Anderson, S.P. and Neven, D.J., 1990, "Spatial competition à la Cournot: price discrimination by quantity-setting oligopolists", *Journal of Regional Science*, 30, No.1, 1-14

Arrow, K.J., 1962, "Economic welfare and the allocation of resources for invention", in Nelson, R.R., (ed.), *The rate and direction of inventive activity: economic and social factors*, Princeton: Princeton University Press, 609-26.

Bagwell, K. and Staiger, R.W., 1992, "The sensitivity of strategic and corrective R&D policy in battles for monopoly", *International Economic Review*, 33, 795-816.

Bagwell, K. and Staiger, R.W., 1994, "The sensitivity of strategic and corrective R&D policy in oligopolistic industries", *Journal of International Economics*, 36, 133-50.

Bain, J.S., 1951, "Relation of profit rate to industry concentration: American manufacturing, 1936-1940", *Quarterly Journal of Economics*, 65, Issue 3, 293-324.

Barnett, V. and Lewis, T., 1984, *Outliers in statistical data*, New York: Wiley.

Beath, J. and Katsoulacos, Y., 1991, *The Economic Theory of Product Differentiation*, Cambridge: Cambridge University Press.

Beaton, A.E. and Tuckey, J.W., 1974, "The fitting of power series, meaning polynomials, Illustrated on band-spectro-scopic data", *Technometrics*, 16, 147-85.

Beckmann, M.J. and Thisse, J.-F., 1986, "The location of production activities", in Nijkamp, P. (ed.), *Handbook of regional and urban economics*, I, Amsterdam: Elsevier Science Publishers Bv., 21-95.

Bernstein, J.I., 1988, "Cost of production, intra- and interindustry R&D spillovers: Canadian evidence", *Canadian Journal of Economics*, XXI, No.2, May, 324-47.

Bernstein, J.I., 1989, "The structure of Canadian inter-industry R&D spillovers, and the rates of return to R&D", *Journal of Industrial Economics*, XXXVII, No.3, March, 315-28.

Bernstein, J.I. and Nadri, M.I., 1988, "Interindustry R&D spillovers, rates of return, and production in high-tech industries", *American Economic Review, Papers and Proceedings*, 78, No.2, May, 429-34.

Bernstein, J.I. and Nadri, M.I., 1989, "Research and development, and intraindustry spillovers: an empirical application of dynamic duality", *Review of Economic Studies*, 56, 249-69.

Bertschek, I., 1995, "Product and process innovations as a response to increasing imports and foreign direct investment", *Journal of Industrial Economics*, XLIII, No.4, December, 341-57.

Bloch, H., 1994, "Sample-selection for estimating the relationship between concentration and profitability from cross-industry data", *Review of Industrial Organization*, 9, 71-84.

Bozeman, B., Link, A. and Zardkoohi, A., 1986, "An economic analysis of R&D joint ventures", *Managerial and Decision Economics*, 7, 263-6.

Brander, J.A. and Spencer, B.J., 1981, "Strategic commitment with R&D: the symmetric case", *Bell Journal of Economics*, 14, No.1, Spring, 225-35.

Bresnahan, F.F. and Schmalensee, R., 1987, "The empirical renaissance in industrial economics: an overview", *Journal of Industrial Economics*, XXXV, No.4, June, 371-8.

Bulow, J., Geanakoplos, J. and Klemperer, P., 1985, "Multimarket oligopoly: strategic substitutes and complements", *Journal of Political Economy*, 93, 488-511.

Caballero, R.J. and Jaffe, A.B., 1993, "How high are the giants' shoulders: an empirical assessment of knowledge spillovers and creative destruction in a model of economic growth", *NBER Working Paper*, No. 4370.

Choi, J.P., 1993, "Cooperative R&D with product market competition", *International Journal of Industrial Organization*, 11, 553-71.

Cohen, W.M. and Levinthal, D.A., 1989, "Innovation and learning: the two faces of R&D", *Economic Journal*, 99, September, 569-96.

Connor, J.M. and Peterson, E.B., 1992, "Market-structure determinants of national brand-private label price differences of manufactured food products", *Journal of Industrial Economics*, XL, No.2, June, 157-71.

Connor, J.M. and Peterson, E.B., 1997, "Market-structure determinants of national brand-private label price differences of manufactured food products: reply", *Journal of Industrial Economics*, XLV, No.2, June, 225-6.

Cook, R.D. and Hawkins, D.M., "Unmasking multivariate outliers and leverage points: comment", *Journal of the American Statistical Association*, 85, No.411, September, 640-6.

Crafts, N., 1992, "Productivity growth reconsidered", *Economic Policy: a European Forum*, No.15, October, 387-414.

Crepon, B. and Duguet,E., 1995, "Research & development, competition and innovation", Cahiers Eco & Maths 95.08, *Université de Paris I (Pantheon-Sorbonne)*.

Dasgupta, R.E. and Masking, E., 1986, "The existence of equilibrium in discontinuous economic games, II: applications", *Review of Economic Studies*, 53, 27-41.

D'Aspremont, C., Gabszewicz, J.J. and Thisse, J.-F., 1979, "On Hotelling's stability in competition", *Econometrica*, 47, No.5, 1145-50.

D'Aspremont, C., Gabszewicz, J.J. and Thisse, J.-F., 1983, "Product differences and prices", *Economics Letters*, 11, 19-23.

D'Aspremont, C., and Jacquemin, A., 1988, "Cooperative and noncooperative R&D in duopoly with spillovers", *American Economic Review*, 78, No.5, December, 1133-7.

D'Aspremont, C., and Jacquemin, A., 1990, "Cooperative and noncooperative R&D in duopoly with spillovers: erratum", *American Economic Review*, 80, No.3, June, 641-2.

Denicolò, V., 1996, "Patent races and optimal patent breadth and length", *Journal of Industrial Economics*, XLIV, No.3, September, 249-65.

De Palma, A., Ginsburgh, V., Papageorgiou, Y.Y. and Thisse, J.-F., 1985, "The principle of minimum differentiation holds under sufficient heterogeneity", *Econometrica*, 53, No.4, 767-81.

Dixit, A., 1986, "Comparative statics for oligopoly", *International Economic Review*, **27**, No.2, February, 107-22.

Eaton, B.C., 1972, "Spatial competition revisited", *Canadian Journal of Economics*, **5**, 268-77.

Economides, N.S., 1984, "The principle of minimum differentiation revisited", *European Economic Review*, **24**, 345-68.

Economides, N.S., 1986, "Minimal and maximal product differentiation in Hotelling's duopoly", *Economics Letters*, **21**, 67-71.

Economides, N.S., 1989, "Symmetric equilibrium existence and optimality in a differentiated product market", *Journal of Economic Theory*, **47**, 178-94.

Edgeworth, F.Y., 1887, "On observations relating to several quantities", *Hermathena*, **6**, 279-85.

European Commission, 1995, "Green paper on innovation", *European Commission*, Brussels and Luxembourg.

European Commission, 1996a, "Information Technologies, Productivity and Employment", *European Commission*, Brussels and Luxembourg.

European Commission, 1996b, "Action plan for innovation", *European Commission*, Brussels and Luxembourg.

Fölster, S., 1995, "Should cooperative R&D be subsidized?: an empirical analysis", in Van Witteloostuijn, A. (ed.), *Market evolution: competition and cooperation*, Dordrecht: Kluwer, 53-68.

Friedman, J.W., 1988, "On the strategic importance of prices versus quantities", *RAND Journal of Economics*, **19**, No.4, Winter, 607-22.

Friedman, J.W. and Thisse, J.-F., 1993, "Partial collusion fosters minimum product differentiation", *RAND Journal of Economics*, **24**, No.4, Winter, 631-45.

Gabszewicz, J.J., Thisse, J.-F., Fujita, M. and Schweizer, U., 1986, *Location Theory*, Chur: Harwood Academic Publishers.

Gabszewicz, J.J., Thisse, J.-F., 1986, "On the nature of competition with differentiated products", *Economic Journal*, **96**, 160-72.

Gabszewicz, J.J., Thisse, J.-F., 1992, "Location", in Aumann, R.J. and Hart, S., (eds.), *Handbook of game theory*, Amsterdam: North Holland.

Galbraith, J.K., 1952, *American capitalism, the concept of countervailing power*, 2nd edition, Oxford: Blackwell (use is made of the reprinted version, 1980).

Gasko, M. and Donoho, D., 1982, "Influential observation in data analysis", *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 104-9.

Geroski, P.A., 1991, *Market dynamics and entry*, Oxford: Blackwell.

Geroski, P.A., 1992, "Vertical relations between firms and industrial policy", *Economic Journal*, **102**, January, 138-47.

Geroski, P.A., 1993, "Antitrust policy towards co-operative R&D ventures", *Oxford Review of Economic Policy*, **9**, No.2, 58-71.

Geroski, P.A., 1996, "Do spillovers undermine the incentive to innovate?", *Conference Proceedings*, the Bureau of Industry Economics.

Geroski, P.A., Machin, S. and Van Reenen, J., 1993, "The profitability of innovating firms", *RAND Journal of Economics*, **24**, No.2, Summer, 198-211.

Gilbert, R.J. and Newbery, D.M.G., 1982, "Preemptive patenting and the persistence of monopoly", *American Economic Review*, 72, No.3, June, 514-26.

Graitson, D., 1982, "Spatial competition à la Hotelling: a selective survey", *Journal of Industrial Economics*, 31, 13-25.

Guzzetti, L., 1995, "A brief history of European Union research policy", *European Commission*, Brussels and Luxembourg.

Hall, B.H. and Mairesse, J., 1995, "Exploring the relationship between R&D and productivity in French manufacturing firms", *Journal of Econometrics*, 65, 263-93.

Hampel, F.R., 1971, "A general qualitative definition of robustness", *Annals of Mathematical Statistics*, 42, 1887-96.

Hampel, F.R., 1974, "The influence curve and its role in robust estimation", *Journal of the American Statistical Association*, 69, 383-93.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986, *Robust statistics. The approach based on influence functions*, New York: Wiley.

Henriques, I., 1990, "Cooperative and noncooperative R&D in duopoly with spillovers: comment", *American Economic Review*, 80, No.3, June, 638-40.

Hinloopen, J., 1994, "Subsidizing cooperative and non-cooperative R&D", Working Paper ECO 94/25, *European University Institute*.

Hinloopen, J., 1995, "Cooperative R&D versus R&D-subsidies: Cournot and Bertrand duopolies", Working Paper ECO 95/26, *European University Institute*.

Hinloopen, J. and Martin, S., 1995, "Comment on estimation and interpretation of empirical studies in industrial economics", Working Paper ECO No. 95/33, *European University Institute*.

Hinloopen, J. and Martin, S., 1997, "Market-structure determinants of national brand-private label price differences of manufactured food products: comment", *Journal of Industrial Economics*, XLV, No.2, June, 219-23.

Hinloopen, J. and Van Marrewijk, J.G.M., 1995, "On the Limits and Possibilities of the Principle of Minimum Differentiation", Working Paper ECO 95/26, *European University Institute*.

Hinloopen, J. and Wagenvoort, J.L.M., 1995, "Robust estimation: an example", Working Paper ECO 95/2, *European University Institute*.

Hinloopen, J. and Wagenvoort, J.L.M., 1997, "On the computation and efficiency of a HBP-GM estimator: some simulation results", *Computational Statistics & Data Analysis*, forthcoming.

Hotelling, H., 1929, "Stability in competition", *Economic Journal*, 39, 41-57.

Huber, P.J., 1964, "Robust estimation of a location parameter", *Annals of Mathematical Statistics*, 35, 73-101.

Jacquemin, A., 1988, "Cooperative agreement in R&D and European antitrust policy", *European Economic Review*, 32, 551-60.

Jaffe, 1986, "Technological opportunity and spillovers of R&D", *American Economic Review*, 76, No.5, December, 984-1001.

Judge, G.C., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T.-C., 1988, *Introduction to the theory and practice of econometrics*", New York: Wiley.

Kamien, M.I., Muller, E. and Zang, I., 1992, "Research joint ventures and R&D cartels", *American Economic Review*, 82, No.5, December, 1293-1306.

Kamien, M.I. and Zang, I., 1993, "Competing research joint ventures", *Journal of Economics & Management Strategy*, **2**, No.1, 23-40.

Kats, A., 1995, "More on Hotelling's stability in competition", *International Journal of Industrial Organization*, **13**, 89-93.

Katz, M.L., 1986, "An analysis of cooperative research and development", *RAND Journal of Economics*, **17**, No.4, Winter, 527-43.

Katz, M.L. and Ordover, J.A., 1990, "R&D cooperation and competition", *Brookings Papers: Microeconomics*, 137-91.

Katz, M.L. and Shapiro, C., 1985, "On the licensing of innovations", *RAND Journal of Economics*, **16**, No.4, Winter, 504-20.

Korres, G.M., 1996, *Technological change and economic growth*, Alderschot: Avebury.

Lerner, A.P. and Singer, H.W., 1937, "Some notes on duopoly and spatial competition", *Journal of Political Economy*, **45**, 145-86.

Mansfield, 1985, "How rapidly does new industrial technology leaks out?", *Journal of Industrial Economics*, **XXXIV**, No.2, December, 217-23.

Mansfield, E., Schwartz, M. and Wagner, S., 1981, "Imitation costs and patents: an empirical study", *Economic Journal*, **91**, December, 907-18.

Mansfield, E., Rapoport, J., Romeo, A., Wagner, S. and Beardsley, G., 1977, "Social and private rates of return from industrial innovations", *Quarterly Journal of Economics*, **XCI**, No.2, May, 221-40.

Mairesse, J. and Hall, B.H., 1996, "Estimating the productivity of research and development: an exploration of GMM methods using data on French and United States manufacturing firms", *NBER Working Paper*, No.5501.

Martin, S., 1993, *Advanced industrial economics*, Oxford: Blackwell.

Martin, S., 1994, "Private and social incentives to form R&D joint ventures", *Review of Industrial Organization*, **9**, 157-71.

Martin, S., 1997, "Public Policies towards cooperation in research and development: the European Union, Japan, the United States", in Comanor, W., Goto, A. and Waverman, L., (eds.), *Competition in a global economy*, London: Routledge, 245-88.

Martin, S., 1996, "R&D joint ventures and tacit product market collusion", *European Journal of Political Economy*, **11**, No.4, April, 733-41.

Mata, J., 1993, "Firm entry and firm growth", *Review of Industrial Organization*, **8**, 567-78.

Motta, M., 1992, "Cooperative R&D and vertical product differentiation", *International Journal of Industrial Organization*, **10**, 643-61.

Mueller, W.F. and Sial, M.H., 1993, "Cyclical variation in the profit-concentration relationship", *Review of Industrial Organization*, **8**, 277-91.

Neven, 1985, "Two stage (perfect) equilibrium in Hotelling's model", *Journal of Industrial Economics*, **XXXIII**, March, 317-25.

Novshek, W., 1980, "Equilibrium in simple spatial (or differentiated product) models", *Journal of Economic Theory*, **22**, 313-26.

Osborne, M.J. and Pitchik, C., 1987, "Equilibrium in Hotelling's model of spatial competition", *Econometrica*, **55**, No.4, 911-22.

Patel, P. and Pavitt, K., 1995, "Patterns of technological activity: their measurement and interpretation", in Stoneman, P. (ed.), *Handbook of the economics of innovation and technological change*, Oxford: Blackwell, 14-51.

Phillips, A., 1976, "A critique of empirical studies of relations between market structure and profitability", *Journal of Industrial Economics*, **XXIV**, No.4, June, 241-9.

Phlips, L., 1996, *Competition policy: a game-theoretic perspective*, Cambridge: Cambridge University Press.

Phlips, L. and Thisse, J.-F., 1982, "Spatial competition and the theory of differentiated markets: an introduction", *Journal of Industrial Economics*, **XXXI**, 1-9.

Plackett, R.L., 1972, "Studies in the history of probability and statistics XXIX: the discovery of the method of least squares", *Biometrika*, **59**, 239-51.

Poyago-Theotoky, J., 1995, "Equilibrium and optimal size of a research joint venture in an oligopoly with spillovers", *Journal of Industrial Economics*, **XLIII**, No.2, June, 209-26.

Rosenkranz, S., 1995, "Innovation and cooperation under vertical product differentiation", *International Journal of Industrial Organization*, **13**, 1-22.

Romano, R.E., 1989, "Aspects of R&D subsidization", *Quarterly Journal of Economics*, **CIV**, Issue 4, November, 863-73.

Rousseeuw, P.J., 1984, "Least median of squares regression", *Journal of the American Statistical Association*, **79**, No.388, December, 871-80.

Rousseeuw, P.J., 1985, "Multivariate estimation with high breakdown point", in Grossman, W., Pflug, G., Vincze, I. and Wertz, W. (eds.), *Mathematical statistics and applications*, **B**, Dordrecht: Reidel Publishing, 283-97.

Rousseeuw, P.J. and Leroy, A.M., 1987, *Robust regression and outlier detection*, New York: Wiley.

Rousseeuw, P.J. and Zomeren, B.C. van, 1990, "Unmasking multivariate outliers and leverage points", *Journal of American Statistical Association*, **85**, No.411, September, 633-39.

Rousseeuw, P.J. and Zomeren, B.C. van, 1991, "Robust distances: simulations and cutoff values", in Stahel, W. and Weisberg, S. (eds.), *Directions in robust statistics and diagnostics, Part II*, New York: Springer Verlag, 195-203.

Rousseeuw, P.J. and Wagner, J., 1994, "Robust regression with a distributed intercept using least median of squares", *Computational Statistics & Data Analysis*, **17**, 65-76.

Salop, C.S., 1979, Monopolistic competition with outside goods, *Bell Journal of Economics*, **10**, 141-56.

Scherer, F.M., 1979, "The welfare economics of product variety: an application to the ready-to-eat cereals industry", *Journal of Industrial Economics*, **XXVIII**, No.2, December, 113-34.

Scherer, F.M., Long, W.F., Martin, S., Mueller, D.C., Pascoe, G., Ravenscraft, D.J., Scott, J.T. and Weiss, L.W., 1987, "The validity of studies with line of business data", *American Economic Review*, **77**, No.1, 205-17.

Schmalensee, R., 1989, "Inter-industry studies of structure and performance", in Scmalensee, R. and Willig, R. (eds.), *Handbook of industrial organization*, **2**, Amsterdam: North Holland.

Shy, O., 1995, *Industrial organization; theory and applications*, Cambridge: MIT Press.

Simpson, D.G., Ruppert, D. and Carroll, R.J., 1992, "On one-step GM estimates and stability of inferences in linear regression", *Journal of the American Statistical Association*, **87**, No.418, June, 439-50.

Simpson, R.D., and Vonortas, N.S., 1994, "Cournot equilibrium with imperfectly appropriable R&D", *Journal of Industrial Economics*, **XLII**, No.1, March, 79-92.

Smithies, A., 1941, "Optimum location in spatial competition", *Journal of Political Economy*, **49**, 423-39.

Spence, M., 1976, "Product differentiation and welfare", *American Economic Review*, **66**, No.2, May, 407-14.

Spence, M., 1984, "Cost reduction, competition, and industry performance", *Econometrica*, **52**, No.1, January, 101-21.

Spencer, B.J. and Brander, J.A., 1983, "International R&D rivalry and industrial strategy", *Review of Economic Studies*, **50**, 707-22.

Steurs, G., 1995, "Inter-industry spillovers: what difference do they make?", *International Journal of Industrial Organization*, **13**, 249-76.

Stigler, S.M., 1981, "Gauss and the invention of least squares", *Annals of Statistics*, **9**, 465-74.

Stoneman, P., 1983, *The economic analysis of technical change*", Oxford: Oxford University Press.

Suzuki, K., 1993, "R&D spillovers and technology transfer among and within vertical keiretsu groups", *International Journal of Industrial Organization*, **11**, 573-91.

Suzumura, K., "Cooperative and noncooperative R&D in an oligopoly with spillovers", *American Economic Review*, **82**, No.5, December, 1307-20.

Varian, H.R., 1984, *Microeconomic analysis*, 2nd edition, New York: Norton.

Vonortas, N.S., 1994, "Inter-Firm cooperation with imperfectly appropriable research", *International Journal of Industrial Organization*, **12**, 413-35.

White, H., 1980, "A heteroscedasticity-consistent covariance estimator and a direct test for heteroscedasticity", *Econometrica*, **48**, No.4, 817-38.

Wildman, S.S., 1984, "A note on measuring surplus attributable to differentiated products", *Journal of Industrial Economics*, **XXXIII**, No.1, September, 123-32.

Yi, S.-S., 1996, "The welfare effects of cooperative R&D in oligopoly with spillovers", *Review of Industrial Organization*, **11**, 681-98.

Ziss, S., 1994, "Strategic R&D with spillovers, collusion and welfare", *Journal of Industrial Economics*, **XLII**, No.4, December, 375-93.