Essays on Competition, Innovation, and Public Policy

Masahito Ambashi

A thesis submitted for the degree of Doctor of Philosophy

Department of Economics

University of Essex

June 2017

This thesis is dedicated

to my wife, Miho, my daughter, Kaho,

and my parents, Hisao and Yoko.

Also, I would like to give special thanks

to Allen Uegawachi for his generous support.

ABSTRACT

Competition and innovation, which comprise the driving force of modern economies, have long been an issue in the economics literature. This thesis mainly highlights these two factors in relation to public policy as applied to various analytical frameworks: (i) technology transfer scheme including a grant-back clause when innovation is cumulative (Chapters 1 and 2); (ii) universities that conduct both research and teaching activities (Chapter 3); and (iii) the relationship between competition and productivity (Chapter 4).

Chapter 1 considers desirable technology transfer in a stream of cumulative innovation. Technology competition is likely to generate social overincentives for innovation. It is demonstrated that a grant-back clause with an appropriate distribution of profits can mitigate social overinvestment in the initial and follow-on technologies.

Chapter 2 analyzes the effect of a grant-back clause on incentives to innovate in accordance with the attributes of innovation: severable (non-infringing) and non-severable (infringing). It is illustrated that a grant-back clause under severable innovation can be socially beneficial because it increases the original licensor's incentive to license.

In Chapter 3, a higher education industry model is examined, where universities conduct research and teaching activities to generate research output and student enrollment. The paradoxical result is that when there is strong substitutability between these two activities, a reduction in not only student enrollment but also research output

can occur in response to an increase in research funds. Additionally, this theoretical analysis is motivated by the empirical challenge using the U.S. higher educational institutions data.

Chapter 4 investigates the causal relationship between the effect of competition and TFP growth based on the Japanese industry-level panel data. It finds that although a positive effect of competition is observable in manufacturing industries, such an effect in non-manufacturing industries may be negative in part of the sample period.

Acknowledgments

First and foremost, I am deeply indebted to my supervisor, Professor Katharine Rockett (University of Essex), for offering her generous time, support, and invaluable feedback in writing the whole part of my thesis. It was Professor Rockett who introduced me to the relatively new field of “innovation economics", which stirred much excitement internally when I was in search for a seed of doctoral research. Since embarking on my research project in this field, she read through all the drafts submitted and provided timely and helpful guidance and constructive suggestions essential to the direction of my research. Most of the ideas embedded in my thesis result from meaningful and fruitful discussions with her. In particular, Chapter 2 is an offspring of our co-authored work led by Professor Rockett and Dr. Pierre Régibeau (Charles River Associates; Imperial College London). Furthermore, Professor Rockett's ongoing support after my departure from the University of Essex to commence work again for the Japanese government has been greatly appreciated.

I would like to thank Dr. Luis Vasconcelos and Professor Simon Weidenholzer (University of Essex) for providing constructive comments and suggestions for Chapters 1 and 2. They both have given me considerable encouragement when there was hesitation on my part in tackling theoretical papers. I believe that without their extended cooperation, the completion of my thesis would not be possible.

In writing Chapter 3, I would like to extend my special thanks to Professor João Santos Silva (University of Surrey) for assuming the role as my supervisor of empirical research. Professor Santos Silva pointed out estimation errors and recommended me to apply a more appropriate empirical method. An introduction to empirical research made by Professor Santos Silva is in large part responsible for my keen interest in microeconometrics and empirical microeconomics. Moreover, I would like to express my appreciation for useful advice particularly offered by Dr. Georg Licht (Centre for European Economic Research, ZEW) and Dr. Andrew Toole (U.S. Patent and Trademark Office) at the Competition and Innovation Summer Seminar 2014 that was held at the International Academy Marmaris.

I would also like to thank Dr. George Symeonidis (University of Essex) and Dr. David Reinstein (University of Exeter) for their detailed comments on Chapter 4. I was also supported by the Research Institute of Economy, Trade and Industry (RIETI) of Japan in writing this paper. I would like to express my gratitude to Professor Kyoji Fukao (Hitotsubashi University) and Dr. Toshiyuki Matsuura (Keio University) for their instructions on the *JIP Database* published by RIETI. I am grateful for helpful comments and suggestions received from Professor Tomohiko Inui (Gakushuin University), who accepted to be the referee of my RIETI Discussion Paper. Moreover, I would like to acknowledge for valuable comments made by Professor Deborah Swenson (University of California Davis), Professor Toshihiro Okubo, and Dr. Hayato Kato (Keio University), which improves the paper to be published from the *Asian Economic Papers* (forthcoming).

I very much appreciate the staff participants and students who have attended seminars, such as the Essex Research Strategy Seminar (RSS), the Competition and Innovation Summer School 2014 Seminar, the RIETI Discussion Paper Seminar, and the Asian Economic Panel Summer 2016. Due to their generous sharing of thoughts and ideas, I was able to further develop and extend my research analytical thinking.

Finally, I wish to mention the financial support from the “long-term overseas study program for government officials" sponsored by the Japanese government, creating an invaluable opportunity to study in a doctoral program.

Contents

ABSTRACT iii

Acknowledgments v

Contents viii

List of Tables xi

List of Figures xiii

Introduction 1

Chapter 1. Technology Competition, Cumulative Innovation, and a Grant-Back Clause 11

1.1. Introduction 11

1.2. Model of technology competition 21

1.3. Equilibrium investment in R&D 28

1.3.1. Research joint venture (RJV) 29

1.3.2. Appropriation without technology transfer 32

1.3.3. License contract without a grant-back clause 36

1.3.4. License contract with a grant-back clause 42

1.3.5. Comparison of social welfare 58

1.4. Consumer surplus with cumulative innovation 64

1.5. Extension: uncertainty in initial innovation 70

1.6. Concluding remarks 76

1.7. Appendices 80

References 86

Chapter 2. Grantbacks and Attributes of Innovation: “But for. . . " Defense of a Grant-Back Clause 90

2.1. Introduction 90

2.2. Model structure 96

2.3. Baseline analysis: one potential licensee 101

2.3.1. Non-severable innovation 101

2.3.2. Severable innovation 105

2.4. Extension 112

2.4.1. Prohibition of territorial restrictions 112

2.4.2. Positive innovation cost 120

2.5. Two heterogeneous licensees 129

2.5.1. Non-severable innovation 130

2.5.2. Severable innovation 144

2.6. Concluding remarks 151

2.7. Appendix: partially prohibited territorial restrictions 154

References 158

Chapter 3. Theory and Empirics of University Research and Teaching: Can We Simultaneously Increase University Research Output and Student Enrollment? 160

3.1. Introduction 160

3.2. Empirical implications 170

3.2.1. Overview of the analysis 170

3.2.2. Summary and discussions 173

3.3. Basic model outline 176

3.3.1. Players 176

3.3.2. Timing of the model 182

3.3.3. Equilibrium solution 182

3.3.4. Comparative statics of a research fund 184

3.4. Modeling of an illustrative case 187

3.4.1. Analysis when substitutability exists 187

3.4.2. Analysis when the degree of substitutability is zero 194

3.4.3. Comparative statics in a simplified setting 198

3.5. Tuition fee is a control variable 205

3.5.1. Analysis of a general case 206

3.5.2. Analysis of an illustrative case 208

3.6. Multiple universities compete for students 219

3.6.1. Competition under fixed tuition fees 221

3.6.2. Competition under controlled tuition fees 226

3.7. Concluding remarks 232

3.8. Appendices 236

3.8.1. Proofs of Propositions and Lemmas 236

3.8.2. Investigation of saddle points 246

3.8.3. Comparative statics of other parameters 247

3.8.4. Increasing returns to scale in a research output function 252

3.8.5. Detailed explanation of an empirical analysis 253

References 272

Chapter 4. Competition Effects on Industrial Productivity: An Analysis of Japanese Industries on the Basis of the Industry-Level Panel Data 276

4.1. Introduction 276

4.2. Survey of existing studies 281

4.2.1. Theoretical backgrounds 281

4.2.2. Empirical studies 283

4.3. Empirical formulation 287

4.4. Dataset 294

4.4.1. Dependent variable (TFP growth) 299

4.4.2. Independent variables and IVs 300

4.4.3. Descriptive statistics 302

4.5. Results 307

4.5.1. Estimation results for the period of 1980-2008 309

4.5.2. Estimation results for the period of 1980-1994 313

4.5.3. Estimation results for the period of 1995-2008 316

4.5.4. Summary and discussions 321

4.6. Concluding remarks 328

4.7. Appendices 331

4.7.1. Industry-specific Lerner index 331

4.7.2. Estimation for industries conducting and not conducting R&D investment 332

4.7.3. First-difference estimation 337

References 341

List of Tables

Table 1.1. Ranking of investment in technologies *R* and *D* (RJV; appropriation without technology transfer; and license contract without a grant-back clause). 40

Table 1.2. The amount of optimal investment in technologies *R* and *D*. 57

Table 1.3. Social welfare when consumer surplus is zero. 59

Table 1.4. Social welfare when consumer surplus is positive ($4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$). 65

Table 2.1. Profits and social welfare (non-severable innovation). 104

Table 2.2. Profits and social welfare (severable innovation). 109

Table 2.3. Attributes of innovation and the use of a grant-back clause. 113

Table 2.4. Profits and social welfare (prohibition of territorial restrictions). 118

Table 2.5. Summary of PCL, PCA, and SC in Case 3. 124

Table 2.6. Profits and social welfare (positive innovation cost). 127

Table 2.7. Royalty rates exchanged between the firms. 132

Table 2.8. Profits and social welfare (NN). 136

Table 2.9. Profits and social welfare (NA). 144

Table 2.10. Profits and social welfare (SN). 148

Table 2.11. Profits and social welfare (SA). 150

Table 2.12. Profits and social welfare (partially prohibited territorial restrictions). 155

Table 3.1. Descriptive statistics of variables. 256

Table 3.2. Total number of doctorates awarded and academic staff in 2011. 258

Table 3.3. Total student enrollment and academic staff in 2011. 259

Table 3.4. Regression analysis of the total number of doctorates awarded. 268

Table 3.5. Regression analysis of the number of science doctorates awarded. 269

Table 3.6. Regression analysis of total student enrollment (1). 270

Table 3.7. Regression analysis of total student enrollment (2). 271

Table 4.1. Summary of variables. 289

Table 4.2. Correlation between independent variables and IVs. 294

Table 4.3. Industry list. 297

Table 4.4. Descriptive statistics. 304

Table 4.5. Estimation of the competition effect for the period of 1980-2008. 312

Table 4.6. Estimation of the competition effect for the period of 1980-1994. 315

Table 4.7. Estimation of the competition effect for the period of 1995-2008. 318

Table 4.8. Estimation of the competition effect for the period of 2000-2008. 320

Table 4.9. Estimation of the competition effect for the period of 1980-1994 (industries conducting and not conducting R&D). 335

Table 4.10. Estimation of the competition effect for the period of 1995-2008 (industries conducting and not conducting R&D). 336

Table 4.11. First-difference estimation of the competition effect. 338

List of Figures

Figure 1.1. Timing of the model. 23

Figure 1.2. Diagram of $R^{O^*}(k)$. 54

Figure 1.3. Diagram of social welfare ($4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$). 66

Figure 2.1. Timing of the model (single licensee). 99

Figure 2.2. Diagram of (c, θ) (severable innovation). 107

Figure 2.3. Diagram of (c, θ) (non-severable innovation; positive innovation cost). 122

Figure 2.4. Diagram of (c, θ) (severable innovation; $k = 1$). 125

Figure 2.5. Diagram of (c, θ) (severable innovation; $k = 2$). 126

Figure 2.6. Timing of the model (multiple licensees). 131

Figure 3.1. U.S. federally funded general R&D versus number of doctorates awarded in 2011. 161

Figure 3.2. U.S. federally funded general R&D versus total student enrollment in 2011. 162

Figure 3.3. Location of a university and students. 189

Figure 3.4. Equilibrium solutions with regard to t classified by F. 197

Figure 3.5. Effect of an increase in F on research and teaching effort. 201

Figure 3.6. Reaction curve of $r(t)$ and the capacity constraint. 210

Figure 3.7. Equilibrium of a large university and a small research institute. 212

Figure 3.8. Comparison of equilibrium variables. 224

Figure 3.9. Equilibrium in teaching competition between small research institutes. 231

Figure 3.10. Diagram of $f(F)$ and $g(F)$. 243

Figure 3.11. Frequency of the total number of doctorates awarded in 2011. 258

Figure 4.1. Average contribution to the real GDP growth rate. 278

Figure 4.2. TFP growth rate of manufacturing and non-manufacturing industries. 278

Introduction

Both competition and innovation are regarded as the main driving forces of modern economies. Whilst the latter has been underscored in the context of endogenous economic growth, the former has been a central issue of economics studies, especially in industrial organization. In essence, competition policies including antitrust law aim at eliminating or correcting abuse caused by imperfect competition or market failure and to make a resource allocation closer to a competitive equilibrium allocation that can prove to be far more efficient (i.e. the first fundamental theorem of welfare economics).

Meanwhile, it has also been expected that competition creates economic dynamism. More precisely, in competitive marketplaces, economic agents (mostly firms, but not limited to them) have strong incentives to win against their competitors, and in that process, they intend to innovate to gain a decisive advantage over others. Innovation is frequently sought as a result of competition, and hence, we can say that "competition is the mother of innovation". On the other hand, there can be a reverse causality, where if innovation has been achieved, it may in turn hinder competition between economic agents due to the existence of intellectual property rights such as patents. Since competition and innovation are reciprocally affected in a complicated manner, policies regarding them are required to be deliberately designed so as to deliver benefits based on an appropriate balance of competition and innovation.

The primary purpose of this thesis is to shed light on the relationship between the effects of competition and incentives for innovation in a specific environment on the basis of solid economic models. The first two chapters provide theoretical models that focus mainly on a grant-back clause, a specific form of technology transfer, when innovation is cumulative. While the first chapter analyzes social welfare under the situation where both licensor (initial innovator) and licensee (follow-on innovator) intend to conduct investment in technology development, the second chapter highlights the attribute of innovation, that is, whether innovation is severable or non-severable. By contrast, the third chapter examines a higher education industry. Supposing a setting where universities conduct both research and teaching activities, this chapter investigates theoretically and empirically how public research funds affect research output and student enrollment generated by universities. Finally, the fourth chapter, utilizing Japanese industry-level semi-macro panel data, analyzes competition effects on industrial productivity (TFP). [1] Rightfully, all four chapters provide unique insights into public policy related to the issues of competition and innovation.

Chapter 1 attempts to investigate what scheme of technology development is desirable in the stream of a cumulative innovation, where an initial technology is an essential tool to innovate a follow-on technology. In general, technology competition is likely to generate social overincentives to invest in technologies when consumer surplus obtained from the whole innovation (i.e. improved final products) is negligible. Because of this, taking into account the balance of incentives for initial and follow-on

[1] The fourth chapter was originally written as my MSc dissertation in economics. After a substantial revision, the former version was submitted to the Research Institute of Economy, Trade and Industry (RIETI) and published as RIETI Discussion Paper Series 13-E-098. For details, see the following URL available at: http://www.rieti.go.jp/jp/publications/dp/13e098.pdf. In addition, this study will be published from the *Asian Economic Papers* (MIT Press Journals) with slight modification.

innovation is of much importance.

Based on a two-firm game-theoretic model, Chapter 1 demonstrates that there is a trade-off between investment in the initial and follow-on technologies when uncertainty exists on the part of a follow-on innovation. In other words, if overinvestment in the follow-on technology is decreased to a socially optimal level, overinvestment in the initial technology becomes exacerbated. In response to this problem, it is revealed that grant-back contracts combined with an appropriate distribution of profits between the licensor and licensee can mitigate social overinvestment in the two technologies, and thereby, social welfare is also improved. This result follows from the feature that grant-back clauses ensure a licensor access to the follow-on technology in return for transferring the initial technology to a licensee. As a consequence, the extreme attractiveness of obtaining an edge in technology competition in both innovation stages is reduced. Among other things, Chapter 1 makes it clear that if the particular desirable distribution of profits as mentioned above is specified by a government authority, socially optimal investment in the initial technology can be attained.

Chapter 1 also shows that competition in the follow-on technology can create higher social welfare as consumer surplus becomes large. This implies that, by increasing the possibility that the improved final products appear in the market, this positive effect of competition may overcome the negative effect of the overincentive problem. Finally, the model is extended to encompass uncertainty on the part of the initial innovation in addition to the follow-on innovation. Hereby, it derives the intuition that the role competition plays in the initial innovation stage can also be of great importance for the advent of improved final products through increasing the possibility of success in the initial innovation, that is, the seed technology.

Chapter 2 investigates the effect of a grant-back clause embedded in a license contract on the incentive to innovate in accordance with attributes of innovation. It is assumed that innovation can be divided into severable (non-infringing) and non-severable (infringing) innovation. A grant-back clause typically requires a licensee to reciprocate further innovation to a licensor without due compensation or with compensation that is not linked to the value of the follow-on innovation. Consequently, competition authorities are inclined to express a concern that grant-back clauses may decrease the licensee's incentive to innovate.

Technology Transfer Guidelines formulated by EU in 2004 on grant-back clauses draws a sharp distinction between severable and non-severable innovation. More precisely, the guideline states that while a grant-back clause applying to non-severable innovation is considered to be innocuous, one applying to severable innovation should be treated with much more skepticism in terms of the incentive of a licensee. However, it is revealed in Chapter 2 that this rule should be open to further debate. The likely defense is that grant-back clauses encourage an original licensor to agree to license her initial technology, because the original licensor may not prefer to license in the first place, as the follow-on improved technology of the licensee is likely to make her initial technology obsolete. To this end, Chapter 2 analyzes not only how the attributes of innovation affect the incentive of a licensor and a licensee to innovate, but also circumstances in which an inclusion of a grant-back clause is justifiable, when the innovation type is severable or non-severable.

By supposing a single licensee at the outset, Chapter 2 demonstrates that a grant-back clause does not change at all the original licensor's incentive to license under non-severable innovation. (She licenses her initial technology regardless of the availability of a grant-back clause.) Contrastingly, it is shown that since a grant-back clause increases the incentive of a licensor under severable innovation, the follow-on innovation that enhances social welfare can be achieved. This result suggests that the "but for" defense with a particular focus on the incentive of a licensor is valid for severable innovation, but not for non-severable innovation, which constitutes a polar opposite to the claim made by the EU. In the light of this contribution, Chapter 2 succeeds in producing a convincing rationale for the "but for" defense under severable innovation in terms of the incentive effect of a grant-back clause on a licensor. Moreover, the "but for" defense of a grant-back clause proves to be sustained for both severable and non-severable innovation when territorial restrictions are prohibited.

Chapter 2 also finds that under severable innovation, a positive innovation cost stifles the licensee's incentive to innovate in addition to the licensor's. With regard to this problem accruing to severable innovation, Chapter 2 indicates that a grant-back clause incentivizes both a licensor and a licensee to achieve the follow-on innovation by affecting an *ex ante* bargaining process between them. Furthermore, extending the model toward an inclusion of two heterogeneous licensees in terms of innovation abilities, it is demonstrated that grant-back clauses can raise the licensor's incentive to undertake the initial innovation and to share it with all the licensees. To sum up, the “but for" defense of a grant-back clause for severable innovation is by and large robust to the variations of the model.

Chapter 3 analyzes a unique higher education industry model, which is distinct from the competition and innovation (technology development) models investigated in Chapters 1 and 2. In Chapter 3, universities obtain research funds from a governmental financing agency and earn tuition revenue from students by establishing a tuition fee. Based on this setting, Chapter 3 examines how mutually-connected research and teaching activities of universities interact to generate research output and student enrollment.

First of all, Chapter 3 intends to derive empirical implications on the basis of U.S. university data in order to motivate the following theoretical analysis. That is, it primarily examines which effects, positive or negative, on research output and student enrollment are more dominant in response to an increase in a research fund. Although the result must be interpreted cautiously, it is demonstrated that research funds allocated to private universities might be ineffective or detrimental to an increase in research output, which is measured by the number of doctorates awarded, possibly due to strong substitutability of research and teaching activities. Moreover, it is found that research funds might negatively affect student enrollment of private universities, which implies that substitutability or the crowding-out effect resulting from small capacities may be in force at this university category.

Next, Chapter 3 argues that substitutability between research and teaching activities is of great significance especially in view of the fact that distribution of research funds affect consequential research output and student enrollment. It is theoretically shown that if substitutability is large enough, not only student enrollment but also research output could decrease in response to an increase in research funds on account of a drastic decline in a total research budget. Although the finding that research funds could have a negative effect on research output is paradoxical, policy makers and university administrative officials should keep this possibility in mind – it results from specific conditions which generate strong substitutability.

Given that the degree of substitutability is zero, the theoretical results totally vary according to whether a tuition fee is fixed or controlled. In the case of a fixed tuition fee, whereas research funds can increase both research output and student enrollment when the capacity of the university to undertake its research and teaching activities is not fully utilized (i.e. a “multiplier effect" is at work), it crowds out student enrollment when the university operates at its full capacity (i.e. a “crowding-out effect" is at work). This simple result comes from the fundamental nature of a university that is evaluated ultimately by research output. Hence, if the university does not have any leeway to work with its capacity, every additional activity is devoted to research.

By contrast, in the case where the tuition fee is controlled by a government authority to maximize tuition revenue, a marginal amount of additional research funds never have any positive effect on student enrollment due to the emergence of a “binary divide" between universities. A “binary divide" means that while a large university operates at its full capacity, a small university does not and optimally opts for lesser activities being incongruous to achieving its maximum potentiality. In these two cases, additional research funds lead only to an increase in research output, but not student enrollment, and especially with regard to a large university, the crowding-out effect operates to the extent of decreasing student enrollment. With this in mind, Chapter 3 maintains that in order to make a small university grow from engaging in marginal activities, it is necessary to provide a sufficient amount of research funds or to enhance the capacity of the university.

Chapter 3 also analyzes the effect of competition among universities that are seeking higher student enrollment, and subsequently compares the results between the cases of single and multiple universities. The analysis of Chapter 3 suggests that in both cases of tuition fee, fixed or controlled, limited competition on teaching activities can lead to an increase in the aggregate student enrollment, assuming that the total amount of research funds is exogenously divided evenly among the multiple universities. On the other hand, it is also demonstrated that in some particular cases, the aggregate research output may be decreased by competition, because the merit of increasing returns to scale in research output is lost.

In the last place, Chapter 4 investigates the causal relationship between the degree of competition, which is measured by the Lerner index, and the TFP growth rate on the basis of the Japanese industry-level panel data (the *JIP Database*) compiled from 1980 to 2008. In existing studies, there has been a wide range of opinion differences on this issue. In other words, whereas it is stated that market competition improves productivity, the Schumpeterian hypothesis claims that market power and large firms are imperative for stimulating innovation.

While production functions are not estimated from firm-level micro data and only industry-level semi-macro data including calculated TFP is employed in this study, an intriguing result is obtained as to the relationship. The central finding of Chapter 4 is that although a positive effect of competition on the TFP growth rate is clearly observable in manufacturing industries throughout the sample period (1980-2008), such an effect in non-manufacturing industries may be slightly negative in the latter half of the sample period (1995-2008). This finding of a negative competition effect may lend support to the claim that the Schumpeterian hypothesis can be applied in the case of non-manufacturing industries. Furthermore, a weak inverted-U shape relation between the competition measure and the TFP growth can be seen to a limited extent in all industries. In view of a policy implication, we must keep in mind that in spite of such a negative relationship between competition and TFP growth in non-manufacturing industries, the hasty conclusion that restricting competition improves their productivities could be misleading.

This thesis as a whole investigates topics regarding competition and innovation based on a variety of settings, with a special focus on public policy. In Chapter 1, new perspectives are obtained regarding incentives for initial and follow-on innovation, and the schemes of technology transfer such as a grant-back clause in the context of cumulative innovation. In Chapter 2, it is proven that a grant-back clause should be defended in order to preserve the incentive of a licensor and to achieve an initial innovation under severable innovation. In Chapter 3, both theoretically and empirically, it is demonstrated that there is a possibility that universities may decrease student enrollment and research output as a consequence of increased research funds. Chapter 4 empirically analyzes the effect of competition on TFP growth, the result of which is totally different between industrial categories.

Chapter 1. Technology Competition, Cumulative Innovation, and a Grant-Back Clause

1.1. Introduction

Economists have had sustained keen interest in the economic mechanism of technological development as a major fuel of economic growth. [1] In the real world, firms constantly undertake research and development (R&D) to obtain competitive advantages by differentiating their products or lowering costs. Innovations generated by R&D activities can also enhance social welfare either by improving the quality of or lowering the price of products. While all these effects of innovations can benefit social welfare, much literature has so far focused on whether R&D incentives of firms are socially optimal: too high or too low. Specifically, the key to the problem is technological appropriability; the degree to which an innovator retains exclusively for the returns to R&D affects the incentive to innovate. [2]

Arrow (1962) points out that technologies cannot be appropriated in nature and that this non-appropriability makes technologies have features of public goods characterized by non-rivalry and non-exclusivity. In other words, an inventor and others can use technologies concurrently (i.e. non-rivalry), and the inventor cannot prevent

[1] Schumpeter (1942) pioneered work on the nature of creative destruction, which he assumed to be the engine of the capitalist economy. Solow (1957) discovered that a large part of economic growth in the U.S. economy (over the period from 1909 to 1949) is attributable to “Solow residuals", and he regarded them as technological progress. Along the lines of Schumpeter’s idea, Romer (1986, 1990), Grossman and Helpman (1991a, 1991b), and Aghion and Howitt (1992) present extensive interest of endogenous growth and innovation.

[2] For example, see Tirole (1989) and Lévêque and Ménière (2004) for a brief survey conducted on technological appropriability, patents, and R&D.

others from using those technologies (i.e. non-exclusivity). Subsequently, Arrow (1962) develops the argument that due to the identical public goods features accruing to technologies, firms may invest in less R&D than normally conducted at the socially optimal level, simply because a majority of firms intend to take a free ride on the technology invented by the other firms. [3]

But technologies can sometimes be appropriated using patent systems, whereby technological appropriability is effectively ensured. [4] A patent is one of property rights that is expected to encourage R&D activities by ensuring for a length of time inventors may appropriate the benefits enabled by their technological development (Nordhaus, 1969). Contrastingly, when firms can appropriate their inventions thorough patent protections, social overinvestment in R&D may arise. It should be noted that only one inventor may generally gain access to exclusive patent rights when the patent system becomes available. [5] This means that other subsequent inventors may infringe the patent right of the first inventor even if they have found successful inventions based on the first innovation. [6] As described later, this somewhat extreme assumption makes a patent race similar to a rank-order tournament, where firms tend

---

[3] Other than technological non-appropriability, R&D investment is likely to be affected by uncertainty concerning GDP growth, policy changes, and non-economic events such as wars, which firms may possibly encounter. Bloom (2014) points to real options and risk aversion of investors as factors that could reduce R&D activities in comparison with a socially optimal level.

[4] In addition to patents, the other reasons why technologies are appropriable in practice are that firms keep their inventions confidential from rivals (i.e. secrecy) and that it takes followers some time to reproduce technologies developed by inventing firms (i.e. lead time advantage). See Cohen et al. (2000), who analyze these factors in a survey of manufacturing industries of the United States (US).

[5] The U.S. and the European Union (EU) adopt a similar patent system, which requires an inventor to demonstrate not only is she the original inventor but also the first filer (first-to-file principle). On the other hand, the Japanese patent law stipulates that a patent right is given to the first inventor (first-to-invent principle).

[6] Actual patent rights are not as strong as assumed in the theoretical model presented later. Hence, this chapter does not consider a “weak patent rights" case.

to invest in R&D beyond a socially optimal level with the intention of being first innovators. As a result, R&D investment conducted by firms other than the winner are socially wasteful (Barzel, 1968). [7]

Meanwhile, if we focus on cumulative innovation, where improvements build on previous advances in the stream of innovations, this picture becomes blurred. Scotchmer (1991, 2004) and Green and Scotchmer (1995) argue that both initial and follow-on innovators' incentives to innovate should be taken into account under cumulative innovation. Scotchmer (1991, 2004) emphasizes that if innovation is separately made by independent innovators so that their externalities of creating further advances are not fully internalized, initial innovators may not have sufficient incentives to diffuse their original technologies into markets. In this connection, as the Coase theorem suggests (Coase, 1960), negotiation involved with technological externalities between initial and follow-on innovators potentially culminates in greater incentives on initial innovators' part to innovate. On this point, we see that licensing practices can strengthen social incentives to innovate first by providing initial innovators with exclusionary property rights that can be traded with follow-on innovators.

In particular, when it comes to a "grant-back clause", it obliges a licensee to grant the right on future advances or improvements in the licensed technology to a licensor of the seed technology (Shapiro, 1985). Since the clause increases the degree of the appropriability attached to an initial innovator, the incentive of a licensor to innovate

[7] Barzel (1968) also assumes that the delay of inventions negatively affects consumer surplus in a dynamic model. Consequently, strictly speaking, the competitive equilibrium in a patent race produces socially overinvestment in R&D only if the effect on consumer surplus is small.

can be preserved and reinforced. On the other hand, the clause is subject to abuse in the sense that a licensor may be able to gain a competitive advantage and establish a dominant position over a licensee. This attractiveness of the first patent holding position is likely to urge firms to be actively engaged in technology competition for initial innovation, and thereby, to provide them with strong incentives to invest in such innovation.

A few previous studies address the effect of a grant-back clause in a license contract on innovative activities. [8] van Dijk (2000) focuses on the social overincentive problem of R&D investment under technology competition considering the situation where an incumbent and a challenger compete in an innovation market. The incumbent has already had an initial technology that can be further innovated by both the incumbent and challenger. Since both firms intend to innovate first (i.e. common pool externalities) and the challenger does not take into account the current profit of the incumbent (i.e. business-stealing externalities), their total R&D investment could be more than the socially optimal level. To this overincentive problem, a grant-back clause plays a role in partially internalizing the common pool externalities. It makes these two firms accept that there is no need to devote excessive efforts to innovative activities, because the incumbent will be entitled to an outcome achieved by the challenger and the challenger has to share an outcome with the incumbent. Hence, van Dijk (2000) lends support for the contract with a grant-back clause on ground that it reduces overinvestment in the innovation toward the socially optimal level compared

[8] For examples of empirical analyses of grant-back clauses, see Leone and Reichstein (2012) and Moreira et al. (2012).

to the license contract in the absence of such a clause.

Choi (2002) analyzes a grant-back contract supposing the situation where a licensor possesses both the core (superior) and peripheral (inferior) technologies. The core technology is assumed to enable the licensee to innovate related products, but it creates a competitor to the licensor. If a complete license contract is not available, and thus, a moral hazard problem occurs to the licensor, an additional payoff should be required to give the licensor a consistent incentive to still transfer the core technology. Choi (2002) asserts that, even in a situation where only the peripheral technology is transferred due to a high royalty cost of the core technology, a grant-back clause can encourage the licensor to transfer the core technology by mitigating the licensor's concern about the entry of a future competitor.

Although van Dijk (2000) and Choi (2002) stress the role of a grant-back clause in solving the overincentive problem of the follow-on innovation stage, they assume that one firm has already achieved an initial innovation. In this sense, the incentive for the entire stream of innovation is not fully investigated by these authors. Unlike these studies, Hatanaka (2012) examines a game-theoretic model where two firms compete for both initial and follow-on technologies. There are two technology competition stages, and the firms conclude a license contract with a grant-back clause between the stages. Hatanaka (2012) maintains that a grant-back clause always increases investment in the initial technology and that it may also increase investment in the follow-on technology. The reason for an increased investment in the initial technology is that grant-back clauses generate a potential benefit of acquiring rights to retain the follow-on innovation. With regard to the follow-on technology, the licensor wants to avoid paying a royalty and the licensee wants to gain more profits by selling it, and hence, follow-on investment may increase. The study conducted by Hatanaka (2012) critically lacks a comprehensive analysis of social welfare, however. As was mentioned before, overinvestment in R&D could pose a serious problem by wasting research resources when several firms compete for technologies. It therefore seems imperative to incorporate an evaluation of social welfare into the model.

This chapter makes an attempt to fill the gaps among these previous studies. It develops a model that sheds light on the nature of both technology competition and cumulative innovation, although the problem of incomplete information contemplated by Choi (2002) is ruled out.

More concretely, a game-theoretic model similar to that of Hatanaka (2012) is employed; two firms determine investment in the initial technology (Stage 1), they conclude a contract of technology transfer (Stage 2), and they determine investment in the follow-on technology (Stage 3). [9] The decisions to invest are made non-cooperatively by the two firms. In view of cumulative innovation, the follow-on technology rests on the initial technology (that is, the former is a “research tool"), and the improved final products appear in the market only when the firms succeed in developing the follow-on technology. A greater part of the analysis posits an uncertainty factor that exists in the follow-on innovation stage for the sake of convenience.

[9] The timing follows Green and Scotchmer (1995) who suppose an *ex-ante* agreement that is reached before firms invest in a second innovation. This *ex-ante* agreement enables follow-on innovators to circumvent sunk costs of their investment, and thus, to resolve a hold-up problem.

Since it is assumed that both technologies are completely patentable and more investment generates a higher probability of achieving technological development first, two firms compete with each other to be the first innovator of these two technologies. [10] It thus follows that only one firm will succeed in innovation, which implies that technological success has perfectly negative correlation across firms. In spite of the different assumption from other conventional models, this rank-order tournament assumption accentuates theoretical results. It is shown that due to these typical characteristics of technology competition, when consumer surplus obtained from technological development is negligible, investment in R&D activities could be excessive compared to the socially optimal level. More precisely, given negligible consumer surplus, technology competition results in undermining the profits of firms through a waste of research resources, and not ever contributes to social welfare via consumer surplus.

In fact, whether actual R&D investment is overly high or overly low largely depends on industrial structures and goods to be developed, associated with the created consumer surplus as well as profits. Now that patent systems exist in most developed countries, firms in high-tech industries have a strong tendency toward seeking patent rights to technologies. [11] We can observe from business surveys that duplicative re-

[10] This model is similar to the one designed by Denicolò (2000) in that technology competition both in the first and second development stages are incorporated. Denicolò (2000) breaks the relationship between the two technologies, so that the value of the second innovation is not dependent upon the first one. (The first technology is not a "research tool".) In addition, he compares various regimes according to the features of the second innovation, such as unpatentable and infringing (UI), patentable and infringing (PI), or patentable and non-infringing (PN). Clearly, this chapter highlights the regime of PI, in which an unauthorized use of the second innovation without a license contract infringes exclusive patent rights to the first innovation. As a result, this chapter is complementary to Denicolò's work.

[11] We can see from "patent trolls" that all R&D activities are not efficiently conducted from the perspective of social welfare. Many of these patents are withheld inside firms and are not utilized. For

search can sometimes occur because of technology competition for exclusive patent rights or dominant market positions. The Ministry of Economy, Trade and Industry (METI) of Japan (2011), which reviewed innovative activities of Japanese firms based on questionnaires, reveals that as many as 61.8% of their R&D activities are considered duplicative with those of competitors in the same industries. In particular, such high-tech industries as chemical (65.8%) and pharmaceutical (64.1%) industries exhibit high duplication, and business managers of these industries recognize that most of their R&D investment is wasted at least in view of their profits. [12] Consequently, the analysis of this chapter would be useful in focusing on such a situation where R&D competition like tournament between firms is prevalent.

Below are brief intuitions that are addressed in this study. In the first place, this chapter's contribution is to analyze the innovation incentives in cumulative innovation under various schemes of technological development: namely, (i) research joint venture (RJV), (ii) appropriation without technology transfer, (iii) license contract without a grant-back clause, and (iv) license contract with a grant-back clause. Although the choice of these technological development schemes by firms is assumed to be exogenously given, a comprehensive analysis in each case is conducted. When consumer surplus is negligible (or simply zero), it is demonstrated that there is generally a trade-off between incentives in a stream of cumulative innovation. More precisely, if the

details, see Bessen et al. (2011) and Cohen et al. (2014), who investigate the economic impact of patent trolls on private and social costs and innovations, respectively.

[12] The representative response in pharmaceutical industries is in what follows: "We feel that R&D activities certainly include duplication within industries. But they would be meaningless unless we become the first innovators to achieve technological development, and therefore, it would be difficult to cooperate with other firms in the phase of product development".

intension to peg investment in the follow-on technology to the socially optimal level by allowing the licensor an exclusive use of the initial technology, overinvestment in the initial technology is almost certain to deteriorate due to the increased attractiveness of this technology.

The focus is then directed to a grant-back clause. It is true that a grant-back clause with a license fee (viz., a licensor can collect a license fee in addition to the right to receive a grant-back) further worsens this incentive trade-off. However, it is proven that a well-drawn up grant-back contract, which encompasses an appropriate distribution of profits (Nash bargaining solution; optimal grant-back contract), can provide a better balance between these innovation streams. That is, not only it decreases investment in the follow-on technology in comparison with the competitive equilibrium level, but also it can greatly reduce overinvestment in the initial technology close to the optimal level. If it is possible to specify a particular distribution of profits benefited from a grant-back contract, overinvestment in the initial technology can be reduced to a point of the socially optimal level. These results extend van Dijk (2000), who analyzes the effect of a grant-back clause solely on the follow-on innovation.

This chapter also analyzes the case where improved final products have significant positive effects on consumer surplus, which seems to be a more plausible setting. Making the assumption that firms cannot extract any consumer surplus, this chapter points out that technology competition for follow-on innovation can enhance social welfare by increasing the probability of improved final products appearing in the market that deliver benefits to consumers. The rank of social welfare varies according to the magnitude of consumer surplus. Specifically, the results show that when consumer surplus is small (large), the grant-back clause involved with an appropriate distribution offers higher (lower) social welfare than other schemes such as a RJV. This is a much broader view than those provided by van Dijk (2000) and Hatanaka (2012), who make no reference to the relation between technological development schemes and change in social welfare in accordance with the magnitude of consumer surplus. By contrast, a policy implication of this study is that there is a need to establish appropriate technological development schemes in tandem with such parameters as technological development costs, uncertainty, and the magnitude of consumer surplus.

Additionally, by introducing an uncertainty factor into an initial innovation stage, this chapter attempts to deduce that technology competition in the initial technology is of major importance for offering improved final products to consumers. It is shown that whereas a RJV cannot be optimal even for any degree of uncertainty in the initial technology, an appropriation without technology transfer is much more resilient to uncertainty, conducting more investment in the initial technology.

The rest of this chapter is organized as follows. Section 1.2 outlines the model of technology competition in a cumulative innovation stream. Section 1.3 analyzes the equilibrium of the model, measures the incentives to innovate both the initial and follow-on technologies, and compares social welfare where consumer surplus is negligible. Section 1.4 explicitly considers significantly positive consumer surplus and highlights a different implication for multi-stage technology competition. Section 1.5 investigates the case where uncertainty is also incorporated in initial innovation. Finally, Section 1.6 concludes followed by appendices and full references.

1.2. Model of technology competition

This chapter constructs a concrete structure of the model with an emphasis on both technology competition and cumulative innovation in investigating how innovation incentives work in each technological development scheme. Despite the specificity of the assumptions, the following simple framework is useful in charting the distinct characteristics of these focuses.

Two firms, denoted by 1 and 2, compete for technologies and a product market. There are two types of technologies: “research" (denoted by R) and “development" (denoted by D). These “research" and “development" technologies correspond to the “initial" and “follow-on" technologies, which were referred to in the previous subsection. Cumulative innovation is assumed to specify that initial technology R is essential for proceeding to the next step of developing follow-on technology D and that technology D generates improved final products. It is also assumed that final products can be improved only thorough the use of technology D. For this reason, even if firms achieve only technology R, final products will not generate any improvement at all. Notably, technology R can be interpreted as a so-called “research tool", which implies that multiple innovative stages are required to produce completed products and the net value of a first innovation alone is nil. [13]

In addition, by postulating that both technologies R and D are completely patentable,

[13] The concept of the research tool model is reviewed by Hall (2007) and Rockett (2010).

the possibility of imitation by rival firms is eliminated. That is, patent breadth is assumed to be so broad that firms find it impossible to invent around original patents. [14] The diversity of patent breadth (as well as length) is excluded from this analysis, but in doing so the nature of the results presented is more stark, which helps to isolate the underlying effects. Moreover, this model does not consider business-stealing externalities that are caused by a challenger who intends to take over the existing market of an incumbent. Alternatively, the common pool externalities, which result from firms innovating first in technology competition, play an important role in each technological development process. As was stated earlier, although the assumptions made here are specific, the model has the advantage of being conspicuous enough that all the relevant effects are clearly viewed.

Timing of the model

The model consists of three stages. Figure 1.1 illustrates the order of timing of the decisions in this extensive-form game.

Stage 1 Firms 1 and 2 choose investment in technology *R*.

Stage 2 The firm which has achieved technology *R* appropriates and does not license it, or discloses it through a license contract with (or without) a grant-back clause.

Stage 3 Firms 1 and 2 choose investment in technology *D*.

[14] O'Donoghue (1998) points to the issue of patent protection in breadth to obtain optimal innovative activities in the quality ladder model, where initial innovation is superseded by follow-on innovation so that every inventor switches places between a leader and a follower.

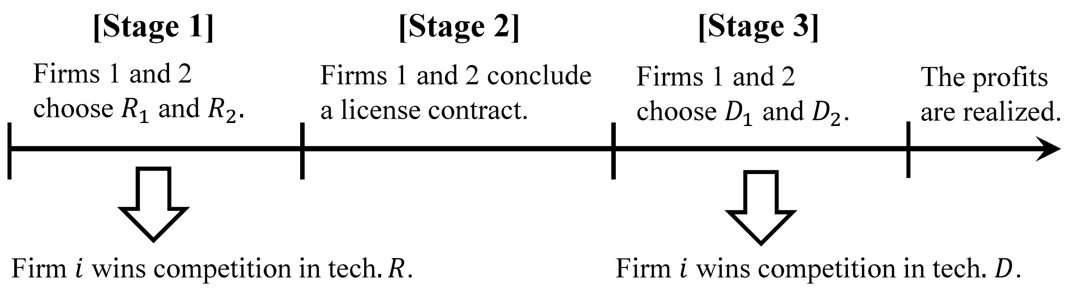


Figure 1.1. Timing of the model.

In Stage 1, the assumption is that when both firms invest in technology R, they cannot achieve it simultaneously. That is, when one firm succeeds in technology R, the other does not. It is posited that if one firm which achieves technology R immediately patents it, the other cannot exploit it without infringing the original patent. This assumption signifies that the initial technology arises as a consequence of entire rank-order tournament led by technology competition, which implies a perfect negative correlation in succeeding innovation across firms.

Let R_i for $i = 1, 2$ denote investment in technology R conducted by firm i. More specifically, following the model established by van Dijk (2000) and Hatanaka (2012), we formulate the probability of firm i achieving technology R such that:

$$P_{R_i}(R_1, R_2) = \frac{R_i}{\sum_{n=1}^{2} R_n} = \frac{R_i}{R_i + R_j}, \quad i, j = 1, 2, \quad i \neq j. \tag{1.1}$$

Equation (1.1) implies that the probability of winning technology competition is determined by the share to total investment. Obviously, if firm i increases its investment, the probability of firm j achieving technology R inevitably declines. In this regard, the firms are typically involved in technology competition seeking an exclusionary right to the initial technology. In addition, as the possibility is excluded that both firms fail to develop technology R, there is no uncertainty in initial innovation. Hereafter, we proceed by supposing that firm 1 realizes technology R without loss of generality.

While this “exclusionary" assumption of technology competition between firms is posited by authors such as van Dijk (2000) and Hatanaka (2012), it is different from other conventional models. In fact, some authors have a preference for employing other formulations of success probabilities. For example, Denicolò (2000) assumes a patent race with a Poisson discovery process (Poisson arrival rate) where the aggregate instantaneous probability of success is simply the sum of the individual probabilities. [15] The advantage of using a Poisson discovery process is generally that; first, it enables us to introduce in a model a cost of delay associated with low investment in technology, and second, it allows for multiple firms simultaneously succeeding in any timing of the stages of the technological development process. However, although acknowledging these advantages, we analyze Equation (1.1) in this chapter to further emphasize the consequences of technology competition.

Next, in Stage 2, there are several cases to be considered. To begin with, firm 1 may withhold technology R, and thus, may not transfer it to firm 2 through a license contract. In this case, firm 2 will be unable to access technology R. As a result, whereas only firm 1 has a chance to develop technology D and to supply improved final products to the market (only if technology D is achieved), firm 2 is forced to

[15] In the patent race model formulated by Denicolò (2000), the payoff function of generic firm i is defined as $\pi_i = \int_0^\infty e^{-(X+r)\tau} x_i v d\tau - c x_i$, where $X = \sum_i x_i$ is aggregate technology investment, r is the discount rate, v is profit from technological development, and c is marginal cost of technology investment. In other words, the payoff of firms is the present value of expected profit, net of technology development cost.

supply existing unimproved products. By contrast, firm 1 may grant full access by disclosing technology R to firm 2 as stipulated in a license contract, which may be required by a government authority, so that firm 2 will be able to utilize technology R by providing a payment to firm 1. Lastly, firm 1 may be able to employ technology D achieved by firm 2 through a license contract with a grant-back clause. Such a clause is expected to weaken the overincentive to develop technology D in the context of technology competition. Note that these technological development schemes are exogenously arranged for firms in this model.

Let D_i denote investment in technology D by firm i in Stage 3. As previously assumed in technology R, technology D can also be appropriated by one firm through a patent immediately following the achievement of the follow-on innovation. Unlike technology R, however, there exists uncertainty about whether technology D can be achieved or not. This assumption is distinct from those of other studies such as Banal-Estañol and Macho-Stadler (2010), who posit uncertainty on the part of initial innovation. The background of our assumption is as follows: while basic innovations have been regarded as difficult to exploit in the natural science fields, it has been recently said that the phase of applications in fields such as pharmaceuticals includes much more uncertainty. [16] Accordingly, this model, among others, analyzes innova-

[16] The process of drug development is split into the two phases: the initial phase of discovering the candidates of chemical compounds that have the potential for new drugs and the follow-on phase of conducting clinical experiments that confirm their usefulness based on human bodies. In general, as the interview results obtained by Saur-Amaral and Borges Gouveia (2007) reveal the difficulty of evaluating uncertainty in clinical experiment phases (i.e. follow-on innovation phases), follow-on innovation seems to display much more uncertainty in a pharmaceuticals case. In surveying the literature on uncertainty in innovations, Jalonen (2012) stresses the importance of classifying the causes of technological uncertainty into a lack of knowledge of new technical details (i.e. basics) and a lack of knowledge required to use this new technology (i.e. applications). Following this classification, our model particularly focuses on the latter cause.

tive activities where a large degree of uncertainty is embedded in follow-on innovation. Meanwhile, we see in Section 1.5 how the new implications are derived if uncertainty exists in the initial innovation stage, too. Introducing an uncertainty factor, $u > 0$, the probability that firm i achieves technology D is formulated as follows (technology competition like rank-order tournament is also assumed):

$$P_{D_i}(D_1, D_2) = \frac{D_i}{\sum_{n=1}^{2} D_n + u} = \frac{D_i}{D_i + D_j + u}, \quad i, j = 1, 2, \quad i \neq j. \tag{1.2}$$

van Dijk (2000) also takes into consideration such an uncertainty factor. Equation (1.2) shows that the larger the u, the smaller the probability that technology D is achieved by either firm. The probability that technology D is not developed by the two firms is represented as $P_{D_u} = \frac{u}{\sum_{n=1}^{2} D_n + u}$, where uncertainty comes together with difficulty or complexity of the follow-on innovation. Consequently, if u gets large, the final stage of innovation becomes extremely difficult goals to achieve given the amount of investment (i.e. as $u \to \infty$, $P_{D_u} = \frac{u}{\sum_{n=1}^{2} D_n + u} \to 0$). In this situation, it is implied that technologies cannot be easily developed through the use of simple automatic algorithms or stereotyped procedures.

In the payoff stage following Stage 3, revenue of final products is realized. If one firm achieves technology D and the other necessarily fails to do (recall the assumption that both firms cannot achieve technology D), the revenue of the former and the latter results in $\overline{\pi}$ and $\underline{\pi}$, respectively, with $\overline{\pi} > \underline{\pi}$. This is quite natural since the successful firm achieving technology D can produce improved final products that result in a competitive advantage over the other. If both firms fail to achieve technology D, they are assumed to obtain the same revenue of π, with $\overline{\pi} > \pi > \underline{\pi}$, by sharing a product market symmetrically.

The third-stage expected profits (i.e. payoffs) of firms 1 and 2 are defined as:

$$V_1 = P_{D_1}\overline{\pi} + P_{D_2}\underline{\pi} + P_{D_u}\pi - \alpha R_1 - \beta D_1 = \frac{D_1\overline{\pi} + D_2\underline{\pi} + u\pi}{\sum_{n=1}^{2} D_n + u} - \alpha R_1 - \beta D_1, \qquad (1.3)$$

$$V_2 = P_{D_2}\overline{\pi} + P_{D_1}\underline{\pi} + P_{D_u}\pi - \alpha R_2 - \beta D_2 = \frac{D_2\overline{\pi} + D_1\underline{\pi} + u\pi}{\sum_{n=1}^{2} D_n + u} - \alpha R_2 - \beta D_2, \qquad (1.4)$$

where $\alpha, \beta > 0$ are common unit costs of developing technologies R and D, respectively, and indicates the efficiency of investment. [17]

In order to facilitate numerical analyses hereafter, we set $\overline{\pi} = 2\pi$ – the revenue obtained from improved final products (2π) is the sum of each revenue obtained from sharing a product market with existing unimproved products (π). This assumption ($\overline{\pi} = 2\pi$) can be justified as follows; if a firm producing improved final products can capture the entire market from its rival and cannot extract any consumer surplus (denoted by C) enhanced by improved final products. [18] (See Section 1.4 for a discussion of how consumer surplus can be interpreted in this model.) Put differently, it is assumed that although the successful firm can seize the whole market opportunity, it cannot exercise any monopoly market power over consumers. The reason for this assumption is that firms cannot always use their dominant position like price discrimination over

[17] Although the constant marginal costs are just one aspect of cost structures, this formulation makes reference to the existing studies such as van Dijk (2000) and Denicolò (2000). Hence, it is implicitly assumed that the firm scales are large enough to avoid increasing cost structures (that is, strictly convex cost functions).

[18] Contrastively, van Dijk (2000) assumes that joint profits of firms exactly equal social welfare. The underlying background of his assumption is, for example, perfect price discrimination is employed by firms competing in the final product market.

consumers because a government authority may enforce regulations with an aim of protecting consumers. Separating profits from consumer surplus in this way provides a keen insight for the possible discrepancy between firms' profits and social welfare. On top of that, $\underline{\pi} = 0$ is also posited in the model.

Based on this simplification, Equations (1.3) and (1.4) are rewritten as:

$$V_1 = \frac{(2D_1 + u)\pi}{D_1 + D_2 + u} - \alpha R_1 - \beta D_1, \tag{1.5}$$

$$V_2 = \frac{(2D_2 + u)\pi}{D_1 + D_2 + u} - \alpha R_2 - \beta D_2. \tag{1.6}$$

From hereon, the model based on Equations (1.5) and (1.6) are investigated.

1.3. Equilibrium investment in R&D

In what follows, this section derives an equilibrium solution of the model. The model is further specified in line with technological development schemes regarding the transfer of technology R. Specifically, the following four cases are separately considered: (i) research joint venture (RJV); (ii) appropriation without technology transfer; (iii) license contract without a grant-back clause; and (iv) license contract with a grant-back clause. By comparing social welfare in the equilibria of these cases, our argument is centered on the desirability of technological development schemes, and on what assumptions the results are dependent. In spite of the exogenous arrangements, this investigation would provide valuable guidance for technology and innovation policies under cumulative innovation. While Section 1.3 assumes that consumer surplus, C, obtained from improved final products is negligible (nearly zero), Section 1.4 regards it significantly positive.

1.3.1. Research joint venture (RJV)

In this subsection, we think about as a benchmark what is socially optimal investment in technology R. One plausible way to internalize negative externalities (common pool externalities) caused by technology competition is to form a single research entity, RJV, at the stage of initial innovation. A RJV, which shares an initial technology within firms as if it was a single entity, can keep its development cost at a minimum.

For analytical purposes, the timing of the game is specified as follows; firms 1 and 2 form a single RJV and achieve technology R (Stage 1), the RJV allows the original two firms to access technology R (Stage 2), and they compete seeking technology D (Stage 3). It is assumed that the RJV discussed here does not persist in Stage 3. [19]

The third-stage expected profits of firms 1 and 2, denoted by V_1^J and V_2^J, are given by Equations (1.5) and (1.6), respectively. We can obtain the following first-order conditions of maximizing V_1^J and V_2^J with regard to investment in technology D:

$$\frac{\partial V_1^J}{\partial D_1^J} = \frac{(2D_2^J + u)\pi}{(D_1^J + D_2^J + u)^2} - \beta = 0, \tag{1.7}$$

$$\frac{\partial V_2^J}{\partial D_2^J} = \frac{(2D_1^J + u)\pi}{(D_1^J + D_2^J + u)^2} - \beta = 0. \tag{1.8}$$

[19] Tao and Wu (1997) and Miyagiwa (2007) theoretically demonstrate that RJVs tend to lead a collusion in a downstream product market. However, such collusion may be rejected by a government authority that has little sympathy toward integrated firms within a full sequence of innovations. The reason is that these firms are more than likely to use their strong market power over other markets including product markets.

From Equations (1.7) and (1.8), $D_i^{J^*} = \frac{\pi-\beta u}{2\beta}$ for $i = 1, 2$ is equilibrium investment. It is demonstrated that the second-order condition is also satisfied. [20] Since $D_i^{J^*}$ should take on a positive value, $D_i^{J^*} > 0$ is needed. [21] This condition is equivalent to $\frac{\pi}{\beta u} > 1$, which means that revenue (π) should be large relative to the marginal cost of developing technology D (β) and the uncertainty factor (u). It is not surprising that given certain revenue, large development costs and uncertainty make it more difficult for firms to invest in that technology. In what follows, we analyze the model by postulating that $\frac{\pi}{\beta u} > 1$ is always satisfied.

Furthermore, it must be the case that $V_i^{J^*} = \frac{(2D_i^*+u)\pi}{D_1^*+D_2^*+u} - \alpha R_i^J - \beta D_i^{J^*} = \frac{\pi+\beta u}{2} - \alpha R_i^J$ exceeds the profit, $V_i^{J^0} = \frac{2\beta u\pi}{\pi+\beta u} - \alpha R_i^J$ that is obtained from $D_i^J = 0$. As is easily shown, because $V_i^{J^*} - V_i^{J^0} = \frac{(\pi-\beta u)^2}{2(\pi+\beta u)} > 0$, we always have $V_i^{J^*} > V_i^{J^0}$ under $\frac{\pi}{\beta u} > 1$. Hence, $D_i^{J^*} = \frac{\pi-\beta u}{2\beta}$ guarantees an equilibrium solution and the corner solution, $D_i^J = 0$, is excluded. [22] Finally, total investment amounts to $D^{J^*} = \sum_{n=1}^{2} D_n^{J^*} = \frac{\pi-\beta u}{\beta}$.

Let us revert to Stage 1, where both firms form a RJV to conduct joint research of technology R. The joint profit of the RJV equals $\Omega^J = \sum_{n=1}^{2} V_n^{J^*} = \pi + \beta u - \alpha R^J$, where $R^J = \sum_{n=1}^{2} R_n^J$. This means that the smaller the investment in technology R, the higher the expected profit of the RJV is. From this, the RJV finds it optimal to cut its investment to the extreme limit, while still maintaining the level that results in technology R being innovated. The interpretation is that only “tackling" the research

---

20 From Equation (1.7), we obtain $\frac{\partial^2 V_1^J}{\partial (D_1^J)^2} = -\frac{2\pi(2D_2^J+u)}{(D_1^J+D_2^J+u)^3} = -2\beta < 0$ for $D_1^{J^*} = D_2^{J^*} = \frac{\pi-\beta u}{2\beta}$. The second-order condition is also the case with Equation (1.8).

21 The negative root $D_i^{J^*} = -\frac{u}{2}$ is also derived, but it cannot be considered as a solution.

22 If both firms collude to conduct $D_i = 0$, the payoff of each firm will be $\pi - \alpha R_i$ and it will exceed $V_i^{J^*} = \frac{\pi+\beta u}{2} - \alpha R_i^J$. However, such collusion is commonly difficult to sustain within the firms based on the reasoning of the “prisoners’ dilemma" in game theory.

process matters in order to innovate technology R and further investment is a mere waste. More generally, the RJV is likely to set $R^{J^*} = \varepsilon > 0$, where ε is minimum research investment that is necessary to achieve technology R. In the current model, ε can be infinitesimally small. [23] [24] It is important to note that this infinitesimal investment does not lose generality if we regard it as essential minimum investment to achieve an initial innovation. [25]

This result suggests that a RJV, analogous to a single entity, is the most conducive innovation system in developing the initial technology when technology competition is subject to a waste in research resources. In other words, since the initial technology is certain to be achieved by any positive investment in the absence of uncertainty, additional research expenditure in that technology is a mere loss of profits. If there are multiple firms competing for the initial technology, the common pool externalities arise, which suggests that investment of multiple firms is totally duplicative. In related literature, for instance, Kamien et al. (1992) point to a similar issue that a RJV cooperating its R&D investment economizes scarce research resources and generates higher profits in a Cournot-type downstream market. Indeed, the above-mentioned METI's survey also reports the responses of Japanese firms, for the example of auto-

[23] It is not natural to suppose that the RJV is allowed to conduct no investment in technology R. Actually, if the RJV chooses $R^J = 0$, follow-on investment in technology D is zero and the RJV obtains $V^J = 2\pi$. However, a RJV, not conducting basic research, contradicts their *raison d'etre.*

[24] The model abstracts from considerations of the duration of investment, unlike the model such as Denicolò (2000). In this analysis, hence, time necessary for undertaking innovation is not related to the amount of investment. Moreover, as regards to robustness for small perturbations, probabilistic factors of investment are eliminated from the model.

[25] It is also feasible to posit that a significant amount of investment, $\underline{R} > \varepsilon$, is required for the achievement of technology R. If we define a minimal required amount of initial investment as significantly positive in this way, there is a need for an implicit assumption that initial investment derived in later cases is also large enough that it always exceeds this threshold, $\underline{R}$.

mobile industries, such that “Firms that have common objectives could ideally form a R&D consortium as a platform from the outset before competing in downstream markets". But we need to keep the point in mind that it may be difficult for private firms to voluntarily form a RJV without the support of public policies, particularly because firms occasionally engage in follow-on technology competition at the end of the RJV formation. (See Subsection 1.3.5 that compares profits.)

It is important that when it comes to profits (or social welfare), we need to take into account the effect of competition on investment in technology D in addition to technology R. As has been shown, investment in technology D is also subject to the competition and the common pool externalities being generated. Subsection 1.3.2 investigates how the negative effect of technology competition in the follow-on innovation can be internalized.

The result that optimal R^{J^*} can be very small is dependent on the assumption that there is no presence of an uncertainty factor in initial innovation. In the meantime, if uncertainty is also assumed to be added on initial innovation, it is possible that technology competition by separate firms can be socially desirable, which is succinctly addressed in Section 1.5.

1.3.2. Appropriation without technology transfer

In turn, this subsection establishes the benchmark of socially optimal investment in technology D. The setting is such that while appropriating technology R, firm 1 does not transfer it to firm 2. On the one hand, only firm 1 has an opportunity to proceed to the follow-on innovation, and on the other hand, firm 2 is not eligible to invest at all in technology D (i.e. $D_2^{A^*} = 0$).

Let us denote V_i^A as the third-stage expected profit of firm i for $i = 1, 2$:

$$V_1^A = \frac{(2D_1^A + u)\pi}{D_1^A + u} - \alpha R_1^A - \beta D_1^A, \tag{1.9}$$

$$V_2^A = \frac{u\pi}{D_1^A + u} - \alpha R_2^A. \tag{1.10}$$

The first-order condition of firm 1 is:

$$\frac{\partial V_1^A}{\partial D_1^A} = \frac{u\pi}{(D_1^A + u)^2} - \beta = 0. \tag{1.11}$$

From Equation (1.11), optimal investment in technology D by firm 1 is provided by $D_1^{A^*} = \sqrt{\frac{u\pi}{\beta}} - u > 0$.[26] Naturally, total investment amounts to $D^{A^*} = \sum_{n=1}^{2} D_n^{A^*} = D_1^{A^*}$.

As $D^{A^*} < D^{J^*}$ can be shown,[27] total investment in technology D is smaller than that of a RJV. In our model currently discussed, an appropriation of the initial technology by a single firm is the best way to mitigate the common pool externalities caused by follow-on technology competition, because it nullifies potential competitive investment by a firm that has not achieved the initial technology. In essence, this optimal investment in the follow-on technology corresponds to that van Dijk (2000) derives, who assumes that a firm which has already possessed the initial technology uses it for the next innovation. The economics of property rights generally indicates

26 $V_1^{A^*} = \frac{(2D_1^{A^*}+u)\pi}{D_1^{A^*}+u} - \alpha R_1^A - \beta D_1^{A^*} = 2\pi + \beta u - 2\sqrt{\beta u\pi} - \alpha R_1^A$ is larger than $V_1^{A^0} = \pi - \alpha R_1^A$ that is obtained from $D_1^A = 0$, because $V^{A^*} - V^{A_1^0} = \pi + \beta u - 2\sqrt{\beta u\pi} = (\sqrt{\pi} - \sqrt{\beta u})^2 > 0 \Leftrightarrow V^{A^*} > V^{A_1^0}$. Hence, firm 1 never opts for $D_1^A = 0$ when it possesses technology R.

27 $D^{J^*} - D^{A^*} = (\frac{\pi-\beta u}{\beta}) - (\sqrt{\frac{u\pi}{\beta}} - u) = \sqrt{\frac{\pi}{\beta}}(\frac{\sqrt{\pi}-\sqrt{\beta u}}{\sqrt{\beta}}) > 0 \Leftrightarrow D^{J^*} > D^{A^*}$ for $\frac{\pi}{\beta u} > 1$.

that resources are used in the most efficient way by a single agent that is responsible for the total benefit. Certainly, even if the number of competing firms increases, the nature of the result that a single firm better internalizes the common pool externalities is the same. As readers may have already recognized, this logic can be also applied to the analysis of technology R discussed in Subsection 1.3.1. Consequently, it is the case that conducting technological development by a single firm avoids a negative effect that technology competition creates. And yet, this result that an appropriation without technology transfer is more desirable for innovating technology D is critically dependent on the assumption that consumer surplus created by improved final products equals zero. If positive consumer surplus obtained from improved final products is explicitly introduced in this model, the current argument should be modified. (We further develop this discussion in Section 1.4.)

Now we consider another possibility that firm 2 wins the competition for acquiring access to technology R. Let us represent $\tilde{V}_1^{A^*}$ as the profit of firm 1 when it fails to achieve technology R. From the symmetry, we can derive $\tilde{V}_1^{A^*} = \frac{u\pi}{\tilde{D}_2^{A^*}+u} - \alpha R_1^A$ with $\tilde{D}_2^{A^*} = D_1^{A^*} = \sqrt{\frac{u\pi}{\beta}} - u$.

Stage 2 (stage of technology transfer) is omitted since the initial technology is appropriated by one firm. The first-stage expected payoff of firm 1, Ω_1^A, is given by:

$$\begin{aligned}\Omega_1^A &= P_{R_1}(R_1^A, R_2^A)V_1^{A^*} + P_{R_2}(R_1^A, R_2^A)\tilde{V}_1^{A^*} \\ &= \frac{R_1^A}{R_1^A + R_2^A}\left[\frac{(2D_1^{A^*} + u)\pi}{D_1^{A^*} + u} - \alpha R_1^A - \beta D_1^{A^*}\right] + \frac{R_2^A}{R_1^A + R_2^A}\left(\frac{u\pi}{D_1^{A^*} + u} - \alpha R_1^A\right). \quad (1.12)\end{aligned}$$

The first-order condition with regard to R_1^A is:

$$\frac{\partial \Omega_1^A}{\partial R_1^A} = \frac{R_2^A}{(R_1^A + R_2^A)^2}\left(\frac{2D_1^{A^*}\pi}{D_1^{A^*} + u} - \beta D_1^{A^*}\right) - \alpha = 0. \tag{1.13}$$

The second-order condition is also satisfied for $\frac{\pi}{\beta u} > 1$. [28] Since $R_1^{A^*} = R_2^{A^*}$ holds in an equilibrium due to the symmetry, [29] equilibrium investment in technology R is:

$$R_1^{A^*} = \frac{1}{4\alpha}\left(\frac{2D_1^{A^*}\pi}{D_1^{A^*} + u} - \beta D_1^{A^*}\right) = \frac{2\pi + \beta u - 3\sqrt{\beta u \pi}}{4\alpha}. \tag{1.14}$$

The assumption of $\frac{\pi}{\beta u} > 1$ ensures that $R_1^{A^*}$ is strictly positive. [30] Total investment in technology R amounts to $R^{A^*} = \sum_{n=1}^{2} R_n^{A^*} = \frac{2\pi+\beta u-3\sqrt{\beta u\pi}}{2\alpha}$. Since $R^{A^*} > R^{J^*}$ generally holds, there is an overinvestment in technology R. The reason is that the firms aspire to seek an exclusionary right to the use of technology R which allows access to developing technology D, and therefore, overinvestment in technology R occurs.

In short, appropriating the initial technology generates a trade-off in the sense that whereas it leads to socially optimal investment in the follow-on technology (technology D), it generates social overinvestment in the initial technology (technology R).

---

28 It can be demonstrated that $\frac{\partial^2 \Omega_1^A}{\partial (R_1^A)^2} = -\frac{2R_2^A}{(R_1^A+R_2^A)^3}\left(\frac{2D_1^{A^*}\pi}{D_1^{A^*}+u} - \beta D_1^{A^*}\right) = -\frac{R_2^A(2\pi+\beta u-3\sqrt{\beta u\pi})}{(R_1^A+R_2^A)^2} = -\frac{R_2^A[(\sqrt{\pi}-\sqrt{\beta u})^2+\sqrt{\pi}(\sqrt{\pi}-\sqrt{\beta u})]}{(R_1^A+R_2^A)^2} < 0$ for $\frac{\pi}{\beta u} > 1$. In later cases, the investigation of the second-order condition has been omitted because we can easily illustrate that it is satisfied.

29 The amount of investment in technology R should be symmetric between firms 1 and 2 from the model setting since both firms are in the same position at the timing of Stage 1. This symmetric feature can simplify the model without loss of generality and is also applied throughout the later analyses.

30 $R_1^{A^*} = \frac{(\sqrt{\pi}-\sqrt{\beta u})^2+\sqrt{\pi}(\sqrt{\pi}-\sqrt{\beta u})}{4\alpha} > 0$ for $\frac{\pi}{\beta u} > 1$.

1.3.3. License contract without a grant-back clause

In this subsection, the investigation points to a case where one firm, achieving technology R, collects a license fee from the other in return for transferring it with an expectation of the licensed firm having achieved technology D. It is necessary to briefly note that this type of a license contract differs from a contract that includes a grant-back clause, in that the former does not include any license provisions regarding the follow-on technology and the latter does. In reality, license contracts without a grant-back clause are frequently observed in licensing practices.

A license contract may not be voluntarily offered by the owner firm of the initial technology unless it benefits from licensing in comparison with sole appropriation. In the current setup, it is quite evident that the owner firm does not have any motive to transfer its initial technology to the other because downstream competition never fails to undermine its expected profit even after payment negotiation that is totally favorable for the firm. In other words, the whole profit is likely to shrink due to overinvestment in the follow-on technology caused by technology competition.

Nevertheless, this chapter assumes that the license contract between these two firms can be forced by a government authority. There are three reasons why we consider such a “compulsory" license contract is required from policy perspectives. First, license contracts can reduce overincentives for initial innovation and can prevent research resources from being wasted. Second, they improve (expected) social welfare by increasing the probability of improved final products appearing in markets through technology competition in the follow-on innovation stage. Third, a licensed firm is expected to enlarge product markets that a licensor cannot attend to individually. [31] Here only the first reason is highlighted and the second and third reasons are not taken into account at present. Section 1.4 analyzes the second reason by positing varying positive consumer surplus (as for the second reason, the degree to which improved final products enhance consumer surplus is now assumed to be zero) and points to the importance of competition in follow-on innovation.

The license fee is assumed to be a fixed amount of f^L rather than a proportional amount to be used. [32] Let us assume that firm 1, possessing essential technology R for follow-on innovation, has a total bargaining power over firm 2, which could be a reasonable presumption. Nevertheless, firm 2 must be guaranteed at least the worst profit that is obtained when technology R is not transferred. That is, we restrict attention in this model to a license fee that is acceptable to a licensee, which can be also the best for a licensor. [33]

Under the assumption that firm 1 achieves technology R, the third-stage expected profits of firms 1 and 2 are defined as:

$$V_1^L = \frac{(2D_1^L + u)\pi}{D_1^L + D_2^L + u} - \alpha R_1^L - \beta D_1^L + f^L, \tag{1.15}$$

[31] If a license contract enables a licensee to exploit new markets that a licensor finds it difficult to reach, the licensor may be able to offer such a contract voluntarily without any government intervention. van Dijk (2000) refers to this licensing feature as a “market enlargement effect". For example, it is possible that a patent holder intends to delegate, through licensing, other firms to serve improved products in a foreign market where she may not gain much profitability due to a geographical barrier.

[32] In a research tool model, distinguishing a license fee between whether it is fixed or proportional to the amount of the usage does not make a significant difference since the initial technology is only a clue to the development of the follow-on technology that is not related to the amount of the use of the initial technology.

[33] This argument could be justified by an implicit political system dominated by firms.

$$V_2^L = \frac{(2D_2^L + u)\pi}{D_1^L + D_2^L + u} - \alpha R_2^L - \beta D_2^L - f^L. \tag{1.16}$$

Since an equilibrium solution in Stage 3 is independent of f^L, it is the same with what was obtained from Equations (1.7) and (1.8): $D_i^{L^*} = D_i^{J^*} = \frac{\pi - \beta u}{2\beta}$ for $i = 1, 2$. Namely, the competitive environment in technology D are common across these two cases. Plugging it into Equations (1.15) and (1.16) equates to $V_1^{L^*} = \frac{\pi+\beta u}{2} - \alpha R_1^L + f^L$ and $V_2^{L^*} = \frac{\pi+\beta u}{2} - \alpha R_2^L - f^L$, respectively.

Turning back to Stage 2, firm 1 transfers technology R, and simultaneously negotiates a license fee, f^L, with firm 2. From Equation (1.10), the minimum third-stage profit of firm 2 is equal to $\hat{V}_2^{A^*} = \frac{u\pi}{D_1^{W^*}+u} - \alpha R_2^L = \sqrt{\beta u \pi} - \alpha R_2^L$. Assuming that firm 2 accepts the license contract that induces the same profit with the appropriation without technology transfer, we can derive f^{L^*} such that:

$$V_2^{L^*} = \hat{V}_2^{A^*} \Rightarrow f^{L^*} = \frac{\pi + \beta u - 2\sqrt{\beta u \pi}}{2} = \frac{(\sqrt{\pi} - \sqrt{\beta u})^2}{2} > 0. \tag{1.17}$$

Finally, in Stage 1, if firm 1 fails to achieve technology R, its expected profit results in $\tilde{V}_1^{L^*} = \pi - \alpha R_1^L - \beta D_1^{L^*} - f^{L^*}$. As a result, firm 1 intends to maximize the following first-stage expected profit with regard to R_1^L:

$$\begin{aligned} \Omega_1^L &= P_{R_1}(R_1^L, R_2^L) V_1^{L^*} + P_{R_2}(R_1^L, R_2^L) \tilde{V}_1^{L^*} \\ &= \pi - \alpha R_1^L - \beta D_1^{L^*} + \left(\frac{R_1^L - R_2^L}{R_1^L + R_2^L} \right) f^{L^*}. \end{aligned} \tag{1.18}$$

The first order-condition is:

$$\frac{\partial \Omega^L}{\partial R_1^L} = \frac{2R_2^L f^{L^*}}{(R_1^L + R_2^L)^2} - \alpha = 0. \tag{1.19}$$

Since $R_1^{L^*} = R_2^{L^*}$ holds in a symmetric equilibrium, we obtain:

$$R_1^{L^*} = R_2^{L^*} = \frac{f^{L^*}}{2\alpha} = \frac{\pi + \beta u - 2\sqrt{\beta u \pi}}{4\alpha} = \frac{(\sqrt{\pi} - \sqrt{\beta u})^2}{4\alpha} > 0. \tag{1.20}$$

Total investment in technology R is exemplified by $R^{L^*} = \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha} > R^{J^*}$.

Lemma 1.1 Comparing total investment in technology R and D between the “appropriation without technology transfer" and the “license contract without a grant-back clause", we obtain:

(1) Total investment in technology R is smaller in the license contract without a grant-back clause than in the appropriation without technology transfer ($R^{L^*} < R^{A^*}$); and

(2) Total investment in technology D is smaller in the appropriation without technology transfer than in the license contract without a grant-back clause ($D^{A^*} < D^{L^*}$).

Lemma 1.1 states that the firms have weaker incentives for technology R and stronger incentives for technology D under the license contract without a grant-back clause than the appropriation without technology transfer. Intuitively, since the firms, including both winners and losers, are guaranteed to utilize the initial technology in Stage 3 through a license contract, the overincentive to achieve the initial technology is weakened. Concurrently, the overincentive is not entirely internalized because the firms still intend to extract a license fee as winners and to avoid paying a costly license fee as losers. By contrast, with regard to technology D, since the firms compete for an exclusionary right to improved final products, total investment is sure to exceed the optimal level. This analysis makes it clear that the follow-on innovation scheme affects initial innovation incentives. As an example, the license contract without a grant-back clause causes a trade-off problem in a multiple innovation stream. But it should be noted that this entire trade-off comes from the assumption that consumer surplus is exactly zero. If consumer surplus significantly increases with improved final products as indicated in Section 1.4, a positive competition effect on follow-on innovation can be critical for an improvement in social welfare.

Lastly, another important point is that when there is no uncertainty in developing technology *R*, the license contract without a grant-back clause is inferior to the RJV that always optimizes (i.e. minimizes) investment in technology *R*. More precisely, while the degree of competition in technology *D* is the same across these two schemes, the RJV undertakes less investment in technology *R*. This relationship is made evident in Table 1.1.

	Optimal	Large	
Technology *R*	$R^{J^*} = \varepsilon$	$R^{L^*} = \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha}$	$R^{A^*} = \frac{2\pi+\beta u-3\sqrt{\beta u\pi}}{2\alpha}$
Technology *D*	$D^{A^*} = \sqrt{\frac{u\pi}{\beta}} - u$	$D^{J^*} = D^{L^*} = \frac{\pi-\beta u}{\beta}$	

Table 1.1. Ranking of investment in technologies *R* and *D* (RJV; appropriation without technology transfer; and license contract without a grant-back clause).

Supplementary note

If a government authority is capable of enforcing a particular license fee of f^L by exercising its regulatory power, socially overinvestment in technology *R* could be further decreased. Since smaller investment in technology *R* is desirable in the current model that assumes no uncertainty in this innovation stage, it is optimal for the government authority to control f^L that achieves $R^{L^*} = \varepsilon$ (where ε is an infinitesimal value). From Equation (1.20), such an f^L is easily calculated as $f^{L^*} \approx 0$.

Lemma 1.2 Suppose that a government authority can control a license fee regarding technology *R* (namely f^L) imposed by a licensor on a licensee. As $f^L \to 0$, the result of a license contract without a grant-back clause is reduced to that derived in the case of a RJV.

It is clear that as a license fee decreases, winning competition for the initial technology becomes less attractive to the firms because the winner is forced to disseminate her technology to the future potential competitor not being covered by sufficient compensation. This is why the firms are motivated to reduce investment in technology *R* to the utmost limit, and then engage in competition seeking technology *D*, which substantially leads to the same result as a RJV derives. It may be, however, difficult for a government authority to implement a particular f^L such as $f^{L^*} \approx 0$, since in general the firms voluntarily determine their license fee based on their relative bargaining positions as assumed in the model. (In this model, licensor, who has already possessed technology *R*, is assumed to have an entire bargaining power over a license fee.) Subsection 1.3.4 examines in detail an optimal distribution of profits to achieve socially desirable initial investment in the context of using a grant-back clause. This chapter still retains the assumption that a government authority cannot control a license fee of a license contract without a grant-back clause.

1.3.4. License contract with a grant-back clause

Two benchmark cases have been discussed, such as a RJV and an appropriation without technology transfer, which lead to optimal investment in technologies R and D, respectively. It has been also demonstrated that a trade-off is produced by a license contract without a grant-back clause that reduces investment in technology R while increasing investment in technology D compared to the appropriation without technology transfer. Taking things one step further, this subsection examines whether a license contract with a grant-back clause resolves or mitigates this trade-off problem.

As was briefly mentioned in Section 1.1, a grant-back clause allows an initial innovator to access the follow-on technology possessed by the other, and thereby, is likely to alter the incentives of firms for the whole sequence of innovation. In analyzing the effect of a grant-back contract, we suppose that although firms are forced to conclude a license contract regarding technology R as in Subsection 1.3.3, a grant-back clause is currently available to a licensor. [34]

Three types of grant-backs are analyzed, where we vary the distribution of (expected) profits obtained from a grant-back clause between the firms. In practice, grant-

[34] The fact of the matter is that grant-back clauses are typically regulated in accordance with attributes of follow-on innovation in countries such as the U.S. and within the EU. Chapter 2 examines the validity of grant-back contracts within a cumulative innovation model associated with attributes of follow-on innovation.

backs take on different forms, and therein, the implications for innovation incentives and social welfare could also vary. Hence, there seems to be for policymakers to have some leeway in establishing how profits should be distributed. First, by analyzing the case where a licensor requires a license fee in return for a license contract regarding technology *R* in addition to the inclusion of a grant-back clause, it is demonstrated that the incentive trade-off between the initial and follow-on technologies is deteriorated. This case has an opposite effect of increasing investment in technology *R*, so that the overincentive problem becomes worse. Second, it is revealed that the Nash bargaining solution, which determines a distribution based on their reservation profits, can be a second-best solution to this trade-off problem. Finally, it is shown that a particular distribution of profits produces optimal investment in technology *R*, if there are any policy tools that enable a government authority to implement this distribution.

Grant-back contract with a license fee (perfect extraction by a licensor)

In the case of a grant-back contract with a license fee, firm 1 (which achieves technology *R*) extracts surplus from firm 2 by imposing a license fee, while firm 2 (which achieves technology *D*) does not require any payment in return for the grant-back of technology *D*. While exclusively possessing the initial technology, firm 1 is assumed to be allowed to collect a license fee in addition to the right to receive a grant-back of the follow-on technology. As regards to the license fee, a fixed amount of $f^G > 0$ is considered. In the sense that firm 2 is not compensated for the follow-on innovation due to the “perfect extraction" of profits by firm 1 as a licensor, the distribution of profits could be biased toward firm 1.

We suppose that in spite of the grant-back clause, firm 2 is still eligible to employ technology D, [35] and therefore, that the both firms equally share the product market earning the equivalent revenue, π, owing to the identical improved final products. This equivalent revenue assumption omits the positive impact of "product market competition" in the presence of improved final products, for example, new market acquisition through an introduction of differentiated goods. More precisely, it is implied that there is no indication of any changes made to the fundamental product market structure on the firm side even after a cumulative innovation is achieved.

The third-stage expected profits, V_1^G and V_2^G, are defined such that:

$$V_1^G = \frac{(2D_1^G + D_2^G + u)\pi}{D_1^G + D_2^G + u} - \alpha R_1^G - \beta D_1^G + f^G, \tag{1.21}$$

$$V_2^G = \frac{(D_2^G + u)\pi}{D_1^G + D_2^G + u} - \alpha R_2^G - \beta D_2^G - f^G. \tag{1.22}$$

The first-order conditions of Equations (1.21) and (1.22) are, respectively:

$$\frac{\partial V_1^G}{\partial D_1^G} = \frac{(D_2^G + u)\pi}{(D_1^G + D_2^G + u)^2} - \beta = 0, \tag{1.23}$$

$$\frac{\partial V_2^G}{\partial D_2^G} = \frac{D_1^G \pi}{(D_1^G + D_2^G + u)^2} - \beta = 0. \tag{1.24}$$

Based on $D_1^G = D_2^G + u$ from Equations (1.23) and (1.24), we obtain $D_1^{G^*} = \frac{\pi}{4\beta}$, $D_2^{G^*} = \frac{\pi - 4\beta u}{4\beta} < D_1^{G^*}$, and $D^{G^*} = \sum_{n=1}^{2} D_n^{G^*} = \frac{\pi - 2\beta u}{2\beta}$. [36] Although $D_1^{G^*} > 0$ always holds,

[35] If a grant-back clause prohibits firm 2 from using its own technology D, firm 2 loses all incentives to invest in technology D, so that the framework would be reduced to an appropriation without technology transfer discussed in Subsection 1.3.2. Antitrust law generally stipulates that license contracts, which totally prohibit licensees from using their improved technologies, are regarded as an unfair trade practice. (See Japan Fair Trade Commission [2016], for example.)

[36] The negative roots are eliminated by a non-negativity condition.

$D_2^{G^*} > 0$ does only for $\frac{\pi}{\beta u} > 4$. [37] On the other hand, $D_2^{G^*} = 0$ holds under $1 < \frac{\pi}{\beta u} < 4$, as $D_2^{G^*}$ does not take on a negative value. For the purpose of subsequent analyses of grant-back clauses, $\frac{\pi}{\beta u} > 4$ is assumed in this subsection, which implies that the grant-back contract needs to come into effect. Notably enough, a grant-back contract leads to smaller investment in technology D of both firms 1 and 2 than a license contract without a grant-back clause, since $D_1^{L^*} - D_1^{G^*} = \frac{\pi - 2\beta u}{4\beta} > 0 \Leftrightarrow D_1^{L^*} > D_1^{G^*}$ and $D_2^{L^*} - D_2^{G^*} = \frac{\pi + 2\beta u}{4\beta} > 0 \Leftrightarrow D_2^{L^*} > D_2^{G^*}$, respectively. Also, substituting $D_1^{G^*} = \frac{\pi}{4\beta}$ and $D_2^{G^*} = \frac{\pi - 4\beta u}{4\beta}$ into Equations (1.21) and (1.22) provides $V_1^{G^*} = \frac{5\pi}{4} - \alpha R_1^G + f^G$ and $V_2^{G^*} = \frac{\pi + 4\beta u}{4} - \alpha R_2^G - f^G$, respectively.

The equilibrium investment, $D_1^{G^*} = \frac{\pi}{4\beta}$ and $D_2^{G^*} = \frac{\pi - 4\beta u}{4\beta}$, is intuitive from the feature of the grant-back. Whereas firm 1's investment in technology D increases the profit of only itself (Equation [1.21]), firm 2's increases the profits of both firms 1 and 2 (Equations [1.21] and [1.22]). Similarly, in terms of the marginal profits, firms 1's marginal expected profit (Equation [1.23]) is larger than firm 2's (Equation [1.24]), assuming for the time being that their investment is symmetric ($D_1^{G^*} = D_2^{G^*}$). For these reasons, we find that firm 1 can afford to invest more in technology D than firm 2.

In Stage 2, likewise the analysis in Subsection 1.3.3, firm 1 is likely to set a license fee, f^G, that ensures the minimum third-stage profit of firm 2. As we have seen, this profit equals $\hat{V}_2^{A^*} = \hat{V}_2^{L^*} = \sqrt{\beta u \pi} - \alpha R_2^G$. Hence, f^G is determined such that:

$$V_2^{G^*} = \hat{V}_2^{A^*} \Rightarrow f^{G^*} = \frac{\pi + 4\beta u - 4\sqrt{\beta u \pi}}{4} = \frac{(\sqrt{\pi} - 2\sqrt{\beta u})^2}{4} > 0. \tag{1.25}$$

[37] It can be also demonstrated that the third-stage profits of firms 1 and 2 obtained from this investment are larger than those obtained from $D_i^G = 0$ for $i = 1, 2$.

Turning back to Stage 1, firms 1 and 2 maximize their profits with regard to R_i^G for $i = 1, 2$. If firm 1 fails to achieve technology R, the profit should be $\tilde{V}_1^{G^*} = \frac{(\tilde{D}_1^{G^*}+u)\pi}{\tilde{D}_1^{G^*}+\tilde{D}_2^{G^*}+u} - \alpha R_1^G - \beta\tilde{D}_1^{G^*} - f^{G^*}$, where $\tilde{D}_1^{G^*} = D_2^{G^*} = \frac{\pi-4\beta u}{4\beta}$ and $\tilde{D}_2^{G^*} = D_1^{G^*} = \frac{\pi}{4\beta}$ due to the symmetry. Consequently, the first-stage expected profit of firm 1 is provided by:

$$\begin{aligned}\Omega_1^G &= P_{R_1}(R_1^G, R_2^G)V_1^{G^*} + P_{R_2}(R_1^G, R_2^G)\tilde{V}_1^{G^*} \\ &= \frac{R_1^G}{R_1^G + R_2^G}\left[\frac{(2D_1^{G^*} + D_2^{G^*} + u)\pi}{D_1^{G^*} + D_2^{G^*} + u} - \alpha R_1^G - \beta D_1^{G^*} + f^{G^*}\right] \\ &+ \frac{R_2^G}{R_1^G + R_2^G}\left[\frac{(\tilde{D}_1^{G^*} + u)\pi}{\tilde{D}_1^{G^*} + \tilde{D}_2^{G^*} + u} - \alpha R_1^G - \beta\tilde{D}_1^{G^*} - f^{G^*}\right]. \qquad (1.26)\end{aligned}$$

The first-order condition is:

$$\frac{\partial\Omega_1^G}{\partial R_1^G} = \frac{R_2^G}{(R_1^G + R_2^G)^2}\left[\frac{2D_1^{G^*}\pi}{D_1^{G^*} + D_2^{G^*} + u} - \beta(D_1^{G^*} - D_2^{G^*}) + 2f^{G^*}\right] - \alpha = 0. \qquad (1.27)$$

In order to derive Equation (1.27), $D_1^{G^*} = \tilde{D}_2^{G^*}$ and $D_2^{G^*} = \tilde{D}_1^{G^*}$ have been used. Since $R_1^{G^*} = R_2^{G^*}$ holds in an equilibrium, below is equilibrium investment in technology R:

$$R_1^{G^*} = R_2^{G^*} = \frac{\pi - \beta u + 2f^{G^*}}{4\alpha} = \frac{3\pi + 2\beta u - 4\sqrt{\beta u\pi}}{8\alpha} > 0. \qquad (1.28)$$

Total investment in technology R amounts to $R^{G^*} = \frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha}$. We can see that revenue (π) positively affects investment in technology R while the marginal costs of developing the both initial and follow-on technologies (α and β) and the uncertainty factor (u) negatively affect it. [38]

Proposition 1.1 compares the amounts of investment between a license contract

[38] It is clear that α negatively affects R^{G^*}. With regard to the effect of β on R^{G^*}, we can derive $\frac{\partial R^{G^*}}{\partial\beta} = \frac{2\sqrt{u}(\sqrt{\beta u}-\sqrt{\pi})}{\sqrt{\beta}} < 0$ for $\frac{\pi}{\beta u} > 1$. Moreover, from the symmetric characteristics, $\frac{\partial R^{G^*}}{\partial u} < 0$ is obvious.

that includes a grant-back clause associated with a license fee and the other schemes analyzed up to this point.

Proposition 1.1 With regard to total investment in technologies R and D, we obtain:
(1) $R^{A^*} > R^{G^*} > R^{L^*} > R^{J^*}$ for investment in technology R; and
(2) $D^{L^*} = D^{J^*} > D^{G^*} > D^{A^*}$ for investment in technology D.

By the use of a grant-back clause, overinvestment in technology D is mitigated compared to a license contract without a grant-back clause. There are two reasons for this. The first reason is that the licensor who holds an exclusionary right to the initial technology has an expectation to have access to the follow-on technology, too, even if it fails to develop the follow-on technology. The second is that the licensee ends up decreasing her incentive by gaining a lower profit through transferring the follow-on technology to the licensor. All in all, a grant-back clause in a license contract plays a role in decreasing the attractiveness for the two firms to invest in the follow-on technology, so that it can partially internalize overinvestment.

By contrast, although this type of a grant-back clause is sure to reduce the over-incentive to develop technology R compared to an appropriation without technology transfer, the degree of internalization is less than that of a license contract without a grant-back clause. Since the firms are expected to receive a grant-back of technology D without any payment to the other when they are a licensor of technology R, winning technology competition at the initial innovation stage and retaining the right to receive a grant-back is rather attractive. This “reward" to a licensor is more than a license contract without a grant-back clause, and hence, the observation is made that the overincentive to invest in technology *R* is further exacerbated.

van Dijk (2000) points out that a license contract with a grant-back clause may lead to smaller investment in the follow-on technology by partially internalizing the common pool externalities as the economic mechanism mentioned above, but this model here derives an implication for the initial innovation, too. While van Dijk (2000) does not provide any reference to implications for the initial technology, this analysis explicitly investigates it. Among other things, this chapter reveals that a grant-back clause can internalize the overincentive for the initial technology in comparison with an appropriation without technology transfer, but induces more investment than a license contract without a grant-back clause. However, investment in the initial technology can differ in accordance with the distribution of profits obtained from a grant-back contract. What follows investigates how the incentive to invest in the initial technology alters by introducing different distributions.

Grant-back contract with a Nash bargaining solution

Instead of allowing for a grant-back clause with a license fee, this analysis assumes that a government authority can reinforce firms to negotiate the distribution of (expected) profits obtained from a reduction in overinvestment in the follow-on technology through a grant-back clause. As discussed earlier, since the grant-back contract with a license fee highly biases the profits toward a licensor who has total bargaining power, cooperative negotiation could remedy this bias so as to provide more benefits to a licensee. In this regard, the Nash bargaining solution can be introduced as a way to appropriately distribute the profits. In general, the distribution indicated by the Nash bargaining solution depends on inherent bargaining power among players based on their reservation profits (i.e. “disagreement points"), so that this solution seems to provide one reasonable distribution. [39]

The outside options as disagreement points is assumed to be a license contract without a grant-back clause discussed in Subsection 1.3.3. In other words, the use of a grant-back clause is allowed solely in an environment where the license contract is concluded so as to transfer technology *R*, which is consistent with the assumption set forth in Subsection 1.3.3. This case posits that the conclusion of a license contract regarding the transfer of technology *R* is required by a government authority while a grant-back clause is available to a licensor. Accordingly, when firm 1 achieves technology *R*, the disagreement points of firms 1 and 2 are $\hat{V}_1^{L^*} = \pi + \beta u - \sqrt{\beta u \pi} - \alpha R_1^N$ and $\hat{V}_2^{L^*} = \sqrt{\beta u \pi} - \alpha R_2^N$, respectively. Since the third-stage expected profits, $\hat{V}_1^{G^*} = \frac{5\pi}{4} - \alpha R_1^N$ and $\hat{V}_2^{G^*} = \frac{\pi}{4} + \beta u - \alpha R_2^N$, have been already derived in the previous analyses,

[39] Nash (1950) presents an axiomatic approach, a solution called “Nash bargaining solution", which is derived deductively from an axiomatic system given that agreement with cooperation between players is realized. In addition, Osborne and Rubinstein (1990) examine whether agreement of cooperation is realized by autonomous players without enforcement from the outside. They prove that by introducing an alternating-offer game where players alternately offer a distribution, the subgame perfect equilibrium of this game uniquely exists and that the equilibrium payoffs converge to the Nash bargaining solution as the discount rate becomes close to 1. This chapter does not consider such a non-cooperative approach to the Nash bargaining solution and simply assumes that agreement of cooperation is exercised solely by an outside government authority.

the feasible set of the bargaining is such that: [40]

$$\sum_{n=1}^{2} V_n^N = \sum_{n=1}^{2} \hat{V}_n^{G^*} = \frac{3\pi}{2} + \beta u - \alpha \left(\sum_{n=1}^{2} R_n^N \right)$$

$$\text{subject to} \quad V_1^N \geq \hat{V}_1^{L^*}, \;\; V_2^N \geq \hat{V}_2^{L^*}. \tag{1.29}$$

Assuming that the firms conclude a license contract with a grant-back clause in Stage 2, we can derive the Nash bargaining solution of expected profits from the following problem:

$$\max_{V_1^N, V_2^N} (V_1^N - V_1^{L^*})(V_2^N - V_2^{L^*})$$

$$\text{subject to} \;\; \sum_{n=1}^{2} V_n^N = \sum_{n=1}^{2} \hat{V}_n^{G^*} = \frac{3\pi}{2} + \beta u - \alpha \left(\sum_{n=1}^{2} R_n^N \right). \tag{1.30}$$

Equation (1.30) leads to:

$$V_1^{N^*} = \frac{\sum_{n=1}^{2} \hat{V}_n^{G^*} + (\hat{V}_1^{L^*} - \hat{V}_2^{L^*})}{2} = \frac{5\pi}{4} + \beta u - \sqrt{\beta u \pi} - \alpha R_1^N, \tag{1.31}$$

$$V_2^{N^*} = \frac{\sum_{n=1}^{2} \hat{V}_n^{G^*} + (\hat{V}_2^{L^*} - \hat{V}_1^{L^*})}{2} = \frac{\pi}{4} + \sqrt{\beta u \pi} - \alpha R_2^N. \tag{1.32}$$

Let us denote $\tilde{V}_1^{N^*} = \frac{\pi}{4} + \sqrt{\beta u \pi} - \alpha R_1^N$ as the third-stage expected profit when firm 1 fails to achieve technology R. The first-stage expected profit of firm 1 is:

$$\begin{aligned} \Omega_1^N &= P_{R_1}(R_1^N, R_2^N) V_1^{N^*} + P_{R_2}(R_1^N, R_2^N) \tilde{V}_1^{N^*} \\ &= \frac{R_1^N}{R_1^N + R_2^N} \left(\frac{5\pi}{4} + \beta u - \sqrt{\beta u \pi} - \alpha R_1^N \right) + \frac{R_2^N}{R_1^N + R_2^N} \left(\frac{\pi}{4} + \sqrt{\beta u \pi} - \alpha R_1^N \right). \end{aligned} \tag{1.33}$$

---

[40] An appropriation of technology R is assumed to be unavailable to a licensor. If it is available to a licensing firm as an outside option, the Nash bargaining solution can never be derived because the expected profit of the licensing firm earned by an appropriation without technology transfer always exceeds that realized by a license contract with a grant-back clause. More concretely, it can be demonstrated that $\sum_{n=1}^{2} V_n^N = \frac{3\pi}{2} + \beta u - \alpha(\sum_{n=1}^{2} R_n^N) < \sum_{n=1}^{2} V_n^{A^*} = 2\pi + \beta u - \sqrt{\beta u \pi} - \alpha(\sum_{n=1}^{2} R_n^N)$ for $\frac{\pi}{\beta u} > 4$.

The first-order condition with regard to R_1^N is:

$$\frac{\partial \Omega_1^N}{\partial R_1^N} = \frac{R_2^N}{(R_1^N + R_2^N)^2}(\pi + \beta u - 2\sqrt{\beta u \pi}) - \alpha = 0. \tag{1.34}$$

Using the symmetric characteristics, $R_1^{N^*} = R_2^{N^*}$, an equilibrium solution is obtained:

$$R_i^{N^*} = \frac{\pi + \beta u - 2\sqrt{\beta u \pi}}{4\alpha} = \frac{(\sqrt{\pi} - \sqrt{\beta u})^2}{4\alpha} > 0 \quad \text{for } i = 1, 2. \tag{1.35}$$

$R_i^{N^*} = R_i^{L^*}$ for $i = 1, 2$ should be noted. It is intuitively natural that investment in technology R is equivalent to that generated by a license contract without a grant-back clause, because the “bottom-line" first-stage profits (disagreement points) are also equivalent concerning the two firms. In view of total investment in technology R, we obtain $R^{N^*} = R^{L^*} = \frac{\pi+\beta u-2\sqrt{\beta u \pi}}{2\alpha}$.

Since $R^{L^*} < R^{G^*}$ holds from Proposition 1.1, we can see that the grant-back associated with the Nash bargaining solution reduces overinvestment in the initial technology more than the grant-back clause with a license fee.

Proposition 1.2 $R^{N^*} = R^{L^*} < R^{G^*}$ holds with regard to investment in technology R.

As we have found, overinvestment in technology D can be mitigated by the license contract with a grant-back clause as compared to that without a grant-back clause. Moreover, if the firms are involved with the Nash bargaining process instead of using a grant-back with a license fee through an intervention of a government authority, the overincentive to develop technology R is adjusted to a level that is much closer to the socially optimal level. Thus, the Nash bargaining solution adjusts a distribution of profits, which is overly biased toward a licensor, properly to a licensee. Bearing all these in mind, we can expect that a license contract with a grant-back clause induced by the cooperative negotiation process produces a more desirable result than a license contract without a grant-back clause. In conclusion, given the assumption made so far, it could be justified in a grant-back contract that a government authority requests a bargaining stage to firms, where they negotiate the distribution of profits.

Optimal grant-back contract

In the Nash bargaining solution, total net surplus from the grant-back, $S^* = \sum_{n=1}^{2} \hat{V}_n^{G^*} - \sum_{n=1}^{2} \hat{V}_n^{L^*} = \frac{\pi}{2}$, was equally divided between the two firms. Now suppose that bargaining power is not necessarily equivalent between them. To this end, we define a new parameter $k \in [0, 1]$ as representing a licensor's bargaining power. More precisely, the larger k, the licensor's (or licensee's) bargaining power stronger (weaker). For instance, $k = 1$ and $\frac{1}{2}$ correspond to: a grant-back contract with a license fee (perfect extraction by a licensor) and that with a Nash bargaining solution, respectively.

When firm 1 achieves technology R, the third-stage expected profits are:

$$V_1^{O^*} = \hat{V}_1^{L^*} + kS^* = \frac{(k+2)\pi}{2} + \beta u - \sqrt{\beta u \pi} - \alpha R_1^O, \tag{1.36}$$

$$V_2^{O^*} = \hat{V}_2^{L^*} + (1-k)S^* = \frac{(1-k)\pi}{2} + \sqrt{\beta u \pi} - \alpha R_2^O. \tag{1.37}$$

Let $\tilde{V}_1^{O^*} = \frac{(1-k)\pi}{2} + \sqrt{\beta u \pi} - \alpha R_1^O$ denote the profit when firm 1 fails to achieve technology R. The first-stage expected profit of firm 1 is:

$$\Omega_1^O = P_{R_1}(R_1^O, R_2^O)V_1^{O^*} + P_{R_2}(R_1^O, R_2^O)\tilde{V}_1^{O^*}$$

$$= \frac{R_1^O}{R_1^O + R_2^O}\left[\frac{(k+2)\pi}{2} + \beta u - \sqrt{\beta u \pi} - \alpha R_1^O\right]$$
$$+ \frac{R_2^O}{R_1^O + R_2^O}\left[\frac{(1-k)\pi}{2} + \sqrt{\beta u \pi} - \alpha R_1^O\right]. \tag{1.38}$$

The first-order condition of Equation (1.38) and $R_1^{O^*} = R_2^{O^*}$ indicate:

$$R_i^{O^*}(k) = \frac{(\frac{2k+1}{2})\pi + \beta u - 2\sqrt{\beta u \pi}}{4\alpha} \quad \text{for } i = 1, 2. \tag{1.39}$$

Total investment in technology R amounts to $R^{O^*}(k) = \sum_{n=1}^{2} R_n^{O^*}(k) = \frac{(2k+1)\pi + 2\beta u - 4\sqrt{\beta u \pi}}{4\alpha}$, which is clearly an increasing function of k. In other words, the greater the fraction of the distribution given to a licensor, the greater the investment in technology R.

See Figure 1.2 regarding the diagram of $R^{O^*}(k)$. When $k = 1$, we obviously obtain $R^{O^*}(1) = \frac{3\pi + 2\beta u - 4\sqrt{\beta u \pi}}{4\alpha} = R^{G^*}$. The reason for the coincidence is that when a licensor makes the licensee's profit decrease to the minimum level (an appropriation without technology transfer or a license contract without a grant-back clause) by extracting all surplus through a license fee, the licensor's profit increases by this amount, and hence, the profit maximization problem becomes the same as that of Ω_1^G (Equation [1.26]).

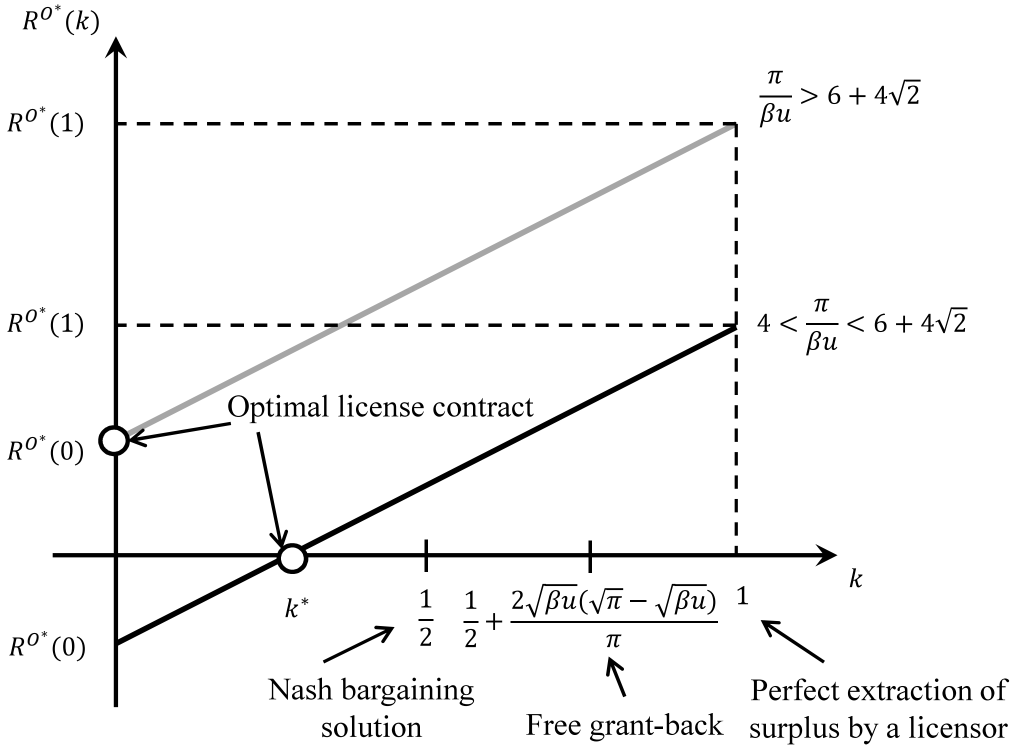


Figure 1.2. Diagram of $R^{O^*}(k)$.

Next, let us denote R^{F^*} as investment in technology R when neither a licensor nor a licensee requires a payment in return for the transfer of technology R and the grant-back of technology D. (This type of a grant-back contract can be termed a "free grant-back".) By plugging $f^{G^*} = 0$ into Equation (1.28), we can obtain $R^{F^*} = \frac{\pi - \beta u}{2\alpha} > 0$. From this relation, $R^{O^*}(\frac{1}{2} + \frac{2\sqrt{\beta u}(\sqrt{\pi} - \sqrt{\beta u})}{\pi}) = R^{F^*}$ can be derived. (Namely, $k = \frac{1}{2} + \frac{2\sqrt{\beta u}(\sqrt{\pi} - \sqrt{\beta u})}{\pi}$.) [41] Importantly, since $\frac{1}{2} + \frac{2\sqrt{\beta u}(\sqrt{\pi} - \sqrt{\beta u})}{\pi} > \frac{1}{2}$, we see that this free grant-back still attaches a greater distribution to a licensor than the Nash bargaining solution. This suggests that the Nash bargaining solution induces a licensor to compensate a licensee who grants technology D to the licensor, which makes an actual payment flow from the licensor to the licensee. Moreover, we always have

[41] $k = \frac{1}{2} + \frac{2\sqrt{\beta u}(\sqrt{\pi} - \sqrt{\beta u})}{\pi}$ is derived by solving the equation, $R^{O^*}(k) = R^{F^*} \Leftrightarrow \frac{(2k+1)\pi + 2\beta u - 4\sqrt{\beta u \pi}}{4\alpha} = \frac{\pi - \beta u}{2\alpha}$ with regard to k.

$R^{O^*}(1) > R^{O^*}(\frac{1}{2} + \frac{2\sqrt{\beta u}(\sqrt{\pi}-\sqrt{\beta u})}{\pi})$ as predicted, because $1 > \frac{1}{2} + \frac{2\sqrt{\beta u}(\sqrt{\pi}-\sqrt{\beta u})}{\pi}$. This proves that overinvestment in technology R deteriorates by allowing a licensor to collect a maximum license fee with a total bargaining power. As we have seen, it is not desirable in the present setting to give a licensor a right to demand the entire distribution.

From these discussions, we can find a distribution that provides optimal investment in technology R. Proposition 1.3 summarizes the result.

Proposition 1.3 (1) If it is possible to specify a particular distribution of expected profits obtained from concluding a contract with a grant-back clause, optimal investment in technology R can be achieved ($R^{O^*} = \varepsilon$); and
(2) The Nash bargaining solution ($R^{N^*} = \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha}$) still gives a licensor an overincentive to invest in technology R as compared to the optimal investment level.

The optimal distribution is denoted by approximately $k^{O^*} \approx k^* = \frac{-\pi-2\beta u+4\sqrt{\beta u\pi}}{2\pi} \in (0, \frac{1}{2})$ when $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$ ($\simeq$ 11.657) holds. [42] Consequently, if the both firms agree with this optimal distribution, investment in technology R can be set at the optimal level, $R^{O^*}(k^{O^*}) = \varepsilon$. ($\varepsilon$ is an infinitesimal positive value.) On the other hand, when the revenue, π, is large enough to hold $\frac{\pi}{\beta u} > 6 + 4\sqrt{2}$ ($\simeq$ 11.657) > 4, it is optimal to set $k^{O^*} = 0$ ($R^{O^*}(0) = \frac{\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha} > 0$), where a licensee should be entitled to receive the entire distribution.

Table 1.2 summarizes the optimal amount of investment in technologies R and

[42] Both $R^{O^*}(0) < 0$ and $R^{O^*}(\frac{1}{2}) > 0$ constitute $k^{O^*} \in (0, \frac{1}{2})$ from the intermediate-value theorem. The former condition, $R^{O^*}(0) < 0$, provides $R^{O^*}(0) = \frac{\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha} < 0 \Leftrightarrow 6-4\sqrt{2} < \frac{\pi}{\beta u} < 6+4\sqrt{2}$. For a detailed derivation, see the demonstration of Proposition 1.3 in Section 1.7.

D in each technological development scheme. Remarkably, the result of Proposition 1.3 is somewhat analogous to the implication about an *ex-post* license fee regarding follow-on innovation formulated by van Dijk (2000). [43] While an *ex-post* license fee in his model allows for an adjustment of investment in the follow-on technology to the optimal level through controlling a payment flow between a licensor and a licensee, our model proposes in a similar fashion that an *ex-post* arrangement of the appropriate distribution of grant-back profits after initial innovation can lead to optimal incentives to innovate the initial technology. As presented in Table 1.2, the types of grant-backs themselves do not have any effects on investment in the follow-on technology regardless of how the expected grant-back profit is distributed between the firms. But our model signifies that the future distribution can directly affect incentives to invest in the initial technology by changing the attractiveness of exercising a right to receive a grant-back as a licensor. Just as van Dijk (2000) mentions, it is important that optimal investment in the initial technology cannot be achieved solely by the inclusion of a grant-back clause into a license contract, but rather, through a cleverly chosen distribution of future profits.

[43] An *ex-post* license fee in van Dijk (2000)'s model can be either positive or negative. When it is negative, it is paid by a grant-back licensor to a licensee after the licensee has innovated the follow-on technology.

RJV	$D^{J^*} = \frac{\pi-\beta u}{\beta}$	$R^{J^*} = \varepsilon$
Appropriation	$D^{A^*} = \sqrt{\frac{u\pi}{\beta}} - u$	$R^{A^*} = \frac{2\pi+\beta u-3\sqrt{\beta u\pi}}{2\alpha}$
License without GB	$D^{L^*} = \frac{\pi-\beta u}{\beta}$	$R^{L^*} = \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha}$
GB (license fee)	$D^{G^*} = \frac{\pi-2\beta u}{2\beta}$	$R^{G^*} = \frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha}$
GB (NB solution)	$D^{N^*} = \frac{\pi-2\beta u}{2\beta}$	$R^{N^*} = \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha}$
GB (optimal)	$D^{O^*} = \frac{\pi-2\beta u}{2\beta}$	$R^{O^*} = \varepsilon$

Note: 1. Appropriation: appropriation without technology transfer.
2. License without GB: license contract without a grant-back clause.
3. GB (license fee): grant-back contract with a license fee.
4. GB (NB solution): grant-back contract with a Nash bargaining solution.
5. GB (optimal): grant-back contract with an optimal distribution.

Table 1.2. The amount of optimal investment in technologies *R* and *D*.

The above-mentioned argument is important from the perspective of an innovation policy. For example, the European Union (EU) (2004) expresses a serious concern about the accumulation of too strong a position of a licensor on ground that grant-back clauses may impose an unfair trade practice on a licensee. And yet, we discover another precaution for the use of a license contract with a grant-back clause: the over-incentive to achieve the initial technology may be exacerbated through a pronounced patent right attached to a licensor, particularly when a licensor exercises his total bargaining power using a grant-back clause with a license fee. As discussed above, the stronger position of a licensor is sure to attract more investment in the initial technology, and as a result, the common pool externalities are seriously caused (especially when initial innovation does not include uncertainty). Therefore, it seems generally desirable that the benefit attached to a licensor should be adjusted to the optimal level in order to prevent excessive technology competition for initial innovation.

However, it is not guaranteed that the firms successfully conclude a contract with the inclusion of a grant-back clause that specifies optimal investment in the initial technology. One reason is that firms may not be able to reach an *ex-ante* agreement with such a contract before the distribution due to an imperfectly drawn up contract. Another possible reason is that firms can make decisions to choose the Nash bargaining solution when a negotiation process is arranged between them by a government authority. This distribution has a rationale for firms in a sense, but it does not always lead to optimal initial investment as pointed out in Proposition 1.3 (2). In a nutshell, there seems to be some room for government intervention in the dividing-up of the profit obtained from a grant-back contract, although it may be difficult for us to secure an effective policy tool to enforce such a particular distribution system.

1.3.5. Comparison of social welfare

To sum up the discussions in Section 1.3, social welfare is compared in relation to each technological development scheme. Under the premise that consumer surplus obtained from improved final products is zero, the sum of the (expected) firms' profits is simply regarded as social welfare. (With regard to the analysis in the presence of positive consumer surplus, see Section 1.4.) Accordingly, social welfare in each scheme is denoted by $\Omega^{X^*} = \sum_{n=1}^{2} \Omega_n^{X^*}$, with $X = J, A, L, G, N$, and O.

To make a comparison between each scheme, it is assumed that $\frac{\pi}{\beta u} > 4$ is still satisfied, which is required to realize a contract with a grant-back clause. Ω^{O^*} represents social welfare of a grant-back contract that achieves optimal investment in technology R. Moreover, we need to take into account the two cases: that is, $R^{O^*}(k^{O^*}) = \varepsilon$ with $k^{O^*} \approx \frac{-\pi-2\beta u+4\sqrt{\beta u\pi}}{2\pi}$ for $4 < \frac{\pi}{\beta u} < 6+4\sqrt{2}$ and $R^{O^*}(k^{O^*}) = \frac{\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha}$ with $k^{O^*} = 0$ for $\frac{\pi}{\beta u} > 6+4\sqrt{2}$.

Table 1.3 represents social welfare in each technological development scheme. We see that social welfare does not depend on the marginal development cost of technology R, namely α. Since investment in technology R is reverse proportion to α in every scheme, it is canceled out in the calculation of the profits. This is because symmetric firms can anticipate what will happen in later stages at the timing of Stage 1.

RJV	$\Omega^{J^*} \approx \pi + \beta u$
Appropriation	$\Omega^{A^*} = \frac{2\pi+\beta u+\sqrt{\beta u\pi}}{2}$
License without GB	$\Omega^{L^*} = \frac{\pi+\beta u+2\sqrt{\beta u\pi}}{2}$
GB (license fee)	$\Omega^{G^*} = \frac{3\pi+2\beta u+4\sqrt{\beta u\pi}}{4}$
GB (NB solution)	$\Omega^{N^*} = \frac{2\pi+\beta u+2\sqrt{\beta u\pi}}{2}$
GB (optimal)	
for $4 < \frac{\pi}{\beta u} < 6+4\sqrt{2}$	$\Omega^{O^*} \approx \frac{3\pi+2\beta u}{2}$
for $\beta u > 6+4\sqrt{2}$	$\Omega^{O^*} = \frac{5\pi+2\beta u+4\sqrt{\beta u\pi}}{4}$

Note: 1. Appropriation: appropriation without technology transfer.
2. License without GB: license contract without a grant-back clause.
3. GB (license fee): grant-back contract with a license fee.
4. GB (NB solution): grant-back contract with a Nash bargaining solution.
5. GB (optimal): grant-back contract with an optimal distribution.

Table 1.3. Social welfare when consumer surplus is zero.

Proposition 1.4 When consumer surplus obtained from improved final products is zero, the ranking of social welfare is such that:

(1) $\Omega^{O^*} > \Omega^{N^*} > \Omega^{A^*} > \Omega^{G^*} > \Omega^{J^*} > \Omega^{L^*}$ for $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$; and

(2) $\Omega^{O^*} > \Omega^{N^*} > \Omega^{A^*} > \Omega^{J^*} > \Omega^{G^*} > \Omega^{L^*}$ for $\frac{\pi}{\beta u} > 6 + 4\sqrt{2}$.

It is necessary to note the assumptions on which Proposition 1.4 relies: we continue to maintain the assumption of the revenue, $\overline{\pi} = 2\pi$ and $\underline{\pi} = 0$, and an uncertainty factor attached to technological development is assumed to be included only in the follow-on technology. Given these assumptions, the ranking of social welfare is determined depending solely on parameters, $\frac{\pi}{\beta u}$, and most of the ranking is the same between (1) and (2) of Proposition 1.4.

A license contract with a grant-back clause, excluding the case of a grant-back contract with a license fee, leads to higher social welfare than the other schemes do. There are two reasons for this result in terms of innovation incentives. The first reason is that grant-back contracts mitigate the overincentive to innovate the follow-on technology through expectations of sharing it *ex post facto*. The second reason is that if they induce an appropriate distribution of profits (Nash bargaining solution; among other things, optimal distribution) obtained from a grant-back clause between the firms, they reduce the overincentives for the initial technology, too, much closer to the socially optimal level. Specifically, when a particular distribution of profits can be designated, the incentive to innovate an initial technology can be adjusted exactly to the same scale as the socially optimal level.

As we have already noted, while an appropriation without technology transfer reduces overinvestment in the follow-on technology to the optimal level, a considerable amount of overinvestment in the initial technology impairs social welfare. On this point, a grant-back contract associated with an appropriate distribution are much more balanced between the incentives for initial and follow-on innovation. In particular, due to the absence of uncertainty in initial innovation, such a grant-back contract can greatly reduce overinvestment in the initial technology. Consequently, it successfully provides higher social welfare than appropriating the initial technology. It is therefore socially desirable for us to arrange a grant-back contract which is based on an appropriate distribution between a licensor and a licensee, instead of permitting an entire appropriation of the initial technology by a licensor.

On the other hand, we can find that social welfare of a grant-back contract with a license fee is quite low. In fact, this grant-back contract always generates higher expected profits than a contract without a grant-back cause. (The total expected profits of the former always expand compared to the latter). But it significantly increases the attractiveness of winning competition in the initial technology by allowing a licensor to extract all surplus from a licensee, so that a deterioration of the overincentive is the most serious matter. As a result, a grant-back contract with a license fee results in lower social welfare than an appropriation without technology transfer (for relatively large basic revenue, π, it is lower than a RJV).

Whereas a RJV achieves optimal investment in the initial technology, it causes sizable common pool externalities on the follow-on technology and decreases social welfare. This lowered social welfare of a RJV can be attributed to the specific structure to which competition in the both initial and follow-on technologies is relevant. More precisely, the existence of uncertainty attached to follow-on innovation requires the firms to conduct more investment to achieve the follow-on technology, which could largely lower social welfare. Additionally, since competition in the follow-on technology never creates social welfare when there is no consumer surplus, it is a mere waste of research resources. Although investment in the initial technology is also wasted in the other schemes, their investment have values, in the sense that it generates future profits brought about by possessing an exclusive right to the initial technology and reduces potential investment in the follow-on technology. Furthermore, comparing social welfare between a RJV and a license contract without a grant-back clause, the former always provides higher social welfare than the latter. In spite of the same degree of competition observed in the follow-on technology, a RJV saves more research resources alloted to the initial technology.

Supplementary note

This chapter has analyzed so far some different technological development schemes individually: (i) RJV; (ii) appropriation without technology transfer; (iii) license contract without a grant-back clause; and (iv) license contract with a grant-back clause (including grant-back contract with a license fee; grant-back contract with a Nash bargaining solution; and optimal grant-back contract). But it should be noted that not only these technological development schemes are in nature endogenously produced, but also commitment issues arise.

In the first place, it is natural to think that the choice of technological development schemes by firms are affected by their availability. If grant-back contract schemes are not allowed by a government authority, firms are likely to select appropriation without technology transfer from the result of Proposition 1.4. In addition, if the bargaining process between firms are not arranged, grant-back contracts cannot be an optimal choice for them. Thus, it is important for us to simultaneously take into account both prescriptions formulated by a government authority and resultant firms' behaviors regarding technological development, which suggests that technological development schemes themselves are endogenous. Furthermore, it could be that although firms intend to engage in an *ex ante* grant-back clause by backward induction, they may have incentives to hold back those clauses when the first stage has finished. The consideration of such firms' commitment issues requires us to further investigate strategic behaviors firms would take for the choice of technological development schemes.

The noticeable results presented in this chapter are built on the specific assumptions. The direction toward a more realistic setting, mitigating to some extent these assumptions, is such that consumers benefit from improved final products enabled by a series of cumulative innovation. If significantly positive consumer surplus is thus incorporated into the model, the results can vary because a positive competition effect on the follow-on technology would be in turn effectual. This analysis is highlighted in Section 1.4.

1.4. Consumer surplus with cumulative innovation

Section 1.4 probes how improved final products affect the ranking of social welfare in technological development schemes when they generate significantly positive consumer surplus (i.e. $C > 0$). In such a case, competition in the follow-on technology is of great importance since it can raise the probability of improved final products appearing in the market. For this reason, a RJV and a license contract without a grant-back clause, where firms compete to develop the follow-on technology, can create higher social welfare than other schemes.

Presently, we continue to suppose that firms cannot extract any consumer surplus into their profits, which is an important assumption to accentuate the discrepancy between firms' profits and social welfare. By assuming that social welfare (W) equals the sum of firms' profits (Ω) and expected consumer surplus ($\frac{D}{D+u}$ where $D = \sum_{n=1}^{2} D_n$), social welfare is redefined as:

$$W^X = \Omega^X + \frac{D^X C}{D^X + u} \quad \text{with } X = J^*, A^*, L^*, G^*, N^*, \text{and } O^*,$$
$$\text{where } D^X = \sum_{n=1}^{2} D_n^X. \tag{1.40}$$

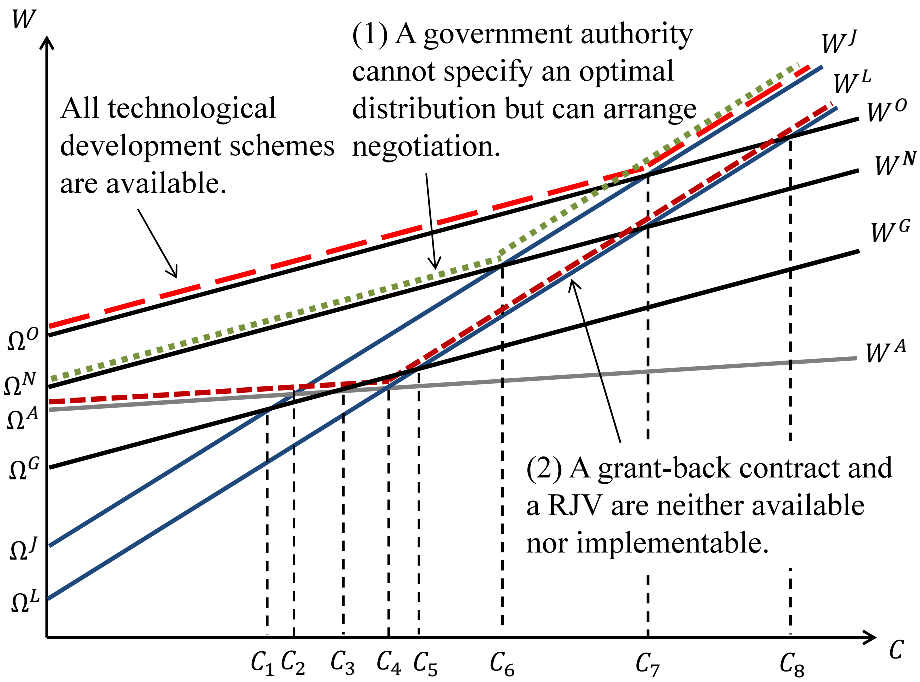
Social welfare is divided up into these two components, which indicates that profit maximization of the firms is not always consistent with social welfare maximization. By separating consumer surplus from profits, the assumption made here is that as consumer surplus changes, profits do not. In other words, this chapter posits that consumer surplus derives from the unique benefit accruing to the use of improved final products resulting from cumulative innovation, which is solely attributable to consumers in the end. (That is, consumer surplus does not derive from a demand expansion or a price change.) Or another assumption that can be made as a special case is that the elasticity of demand for improved final products is exactly zero. This implies that a reduction in the price of products increases demand but does not change the revenue of firms. Taking into consideration the above-mentioned arguments, we can regard consumer surplus formulated in this model as valid. Based on the definition of Equation (1.40), Table 1.4 summarizes social welfare in each technological development scheme by utilizing $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$.

RJV	$W^{J^*} \approx \pi + \beta u + (1 - \frac{\beta u}{\pi})C$
Appropriation	$W^{A^*} = \frac{2\pi+\beta u+\sqrt{\beta u \pi}}{2} + (1 - \sqrt{\frac{\beta u}{\pi}})C$
License without GB	$W^{L^*} = \frac{\pi+\beta u+2\sqrt{\beta u \pi}}{2} + (1 - \frac{\beta u}{\pi})C$
GB (license fee)	$W^{G^*} = \frac{3\pi+2\beta u+4\sqrt{\beta u \pi}}{4} + (1 - \frac{2\beta u}{\pi})C$
GB (NB solution)	$W^{N^*} = \frac{2\pi+\beta u+2\sqrt{\beta u \pi}}{2} + (1 - \frac{2\beta u}{\pi})C$
GB (optimal)	$W^{O^*} \approx \frac{3\pi+2\beta u}{2} + (1 - \frac{2\beta u}{\pi})C$

Note: 1. Appropriation: appropriation without technology transfer.
2. License without GB: license contract without a grant-back clause.
3. GB (license fee): grant-back contract with a license fee.
4. GB (NB solution): grant-back contract with a Nash bargaining solution.
5. GB (optimal): grant-back contract with an optimal distribution.

Table 1.4. Social welfare when consumer surplus is positive ($4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$).

The measurement of social welfare for $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$ is depicted in Figure 1.3 in accordance with the magnitude of consumer surplus. (The discussion in the case of $\frac{\pi}{\beta u} > 6 + 4\sqrt{2}$ is not any different from the below, and hence, it is omitted.) Obviously, the intercepts of the lines correspond to social welfare when consumer surplus is zero, namely, only firms' profits (Ω^{X^*}). It is observed that the slopes of the lines vary in each case; W^{J^*} and W^{L^*} are the steepest, W^{A^*} is the flattest, and W^{G^*}, W^{N^*}, and W^{O^*} are intermediate. Most importantly, the slope of each line is consistent with the intensity of follow-on competition. More precisely, as consumer surplus gets large, fiercer competition in the follow-on technology increases more social welfare through the higher probability of succeeding in the follow-on innovation. The following proposition briefly summarizes the observation obtained from Figure 1.3.



$$C_1 = \frac{\pi(-\pi + 2\beta u + 4\sqrt{\beta u \pi})}{4\beta u} \qquad C_5 = \frac{\pi(2\pi + \beta u + 2\sqrt{\beta u \pi})}{4\beta u}$$

$$C_2 = \frac{\pi}{2} \qquad C_6 = \frac{\pi(2\sqrt{\beta u \pi} - \beta u)}{2\beta u}$$

$$C_3 = \frac{\pi\sqrt{\pi}}{4\sqrt{\beta u}} \qquad C_7 = \frac{\pi^2}{2\beta u}$$

$$C_4 = \frac{\pi\sqrt{\pi}}{2\sqrt{\beta u}} \qquad C_8 = \frac{\pi(2\pi + \beta u - 2\sqrt{\beta u \pi})}{2\beta u}$$

Figure 1.3. Diagram of social welfare ($4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$).

Proposition 1.5 Suppose that positive consumer surplus is created with improved final products as a result of cumulative innovation.

(1) If consumer surplus is relatively small, a grant-back contract associated with an appropriate distribution (Nash bargaining solution; among other things, optimal distribution) is more socially desirable than a RJV (or a license contract without a grant-back clause), and vice versa;

(2) If consumer surplus is relatively large, technological development schemes that firms choose may not induce first-best social welfare; and

(3) A grant-back contract associated with an appropriate distribution is always more socially desirable than an appropriation without technology transfer regardless of the magnitude of consumer surplus.

Proposition 1.5 (1) points to the importance of making various technological development schemes both available and implementable in response to consumer surplus created through improved final products as a result of cumulative innovation.

For example, highlighting a grant-back contract with an optimal distribution and a RJV, we obtain $W^{O^*} > W^{J^*}$ for $C < C_7 = \frac{\pi^2}{2\beta u}$. Intuitively, when consumer surplus obtained from improved final products is relatively small, it is advantageous to save research resources set aside for the follow-on technology using a grant-back contract with an optimal distribution. In general, it is better to employ a grant-back contract than a RJV or a license contract without a grant-back clause in order to save research resources allocated to the follow-on technology. Moreover, it is imperative not only to have a grant-back contract scheme be made available to firms, but also to arrange a negotiation process between them or specify an optimal distribution of profits in order to definitely reduce overinvestment in the initial technology. The firms are likely to choose a technological development scheme from the perspective of their profit maximization (Ω^{O^*}), which is also consistent with social welfare maximization (W^{O^*}), as long as consumer surplus is relatively small ($C < C_7 = \frac{\pi^2}{2\beta u}$).

On the contrary, if consumer surplus is so large that $C > C_7 = \frac{\pi^2}{2\beta u}$, we obtain $W^{J^*} > W^{O^*}$. This implies that when consumer surplus is relatively large, the benefit of the competition effect, increasing the probability of improved final products through a RJV, is expected to be large exceeding the common pool externalities. Nevertheless, as Proposition 1.5 (2) asserts, the firms still prefer a grant-back contract scheme to a RJV, not taking into account consumer surplus because their decisions will be made based solely on the ranking of Ω^{X^*}. Provided that the firms can extract consumer surplus entirely from consumers and that they act as a single entity without competition, it must be the case that their profits totally coincide with social welfare. But in a situation where multiple firms are competing, it is hard to anticipate that profits of firms will be completely equal to social welfare.

In order to achieve maximum social welfare in the above-mentioned case, it is necessary to form a RJV that brings about a strong competition effect in the follow-on technology. Since there is an apparent discrepancy between the profit and social welfare, we may as well entrust a government authority with the power to encourage firms to put in practice a RJV. In this sense, when firms cannot take into account the positive competition effect on (expected) consumer surplus, it is possible to justify policy intervention into technological development schemes to form a RJV. But again, the implementation may also prove to be difficult and costly as it calls for some compensation for firms' involvement with the formation of a RJV, for example, redistribution of consumer surplus to firms.

With regard to Proposition 1.5 (3), even if an arbitrary magnitude of consumer surplus is assumed, $W^{O^*} > W^{N^*} > W^{A^*}$ is always preserved. In other words, although appropriating the initial technology minimizes the cost of technological development in follow-on innovation, its positive effect is, regardless of consumer surplus, smaller than the benefit from a grant-back scheme associated with an appropriate distribution (Nash bargaining solution and optimal distribution). This result is not surprising, since we have already derived $V^{O^*} > V^{N^*} > V^{A^*}$ and the positive competition effect of grant-backs is necessarily greater than an appropriation without technology transfer as consumer surplus is increased. Hence, although enforceability is still open for debate, it might be reasonable to prohibit the use of the appropriation when uncertainty is accrued solely to the follow-on technology. [44]

In conclusion, while the existing studies (van Dijk, 2000; Hatanaka, 2012) have not considered innovation features such as enhancement of consumer surplus separated from profit in their welfare analyses simply by assuming that consumer surplus is zero or social welfare is regarded as firms' profits, our analysis sheds new light on the aspect of choosing appropriate technological development schemes in accordance with the magnitude of consumer surplus.

[44] Repeatedly, we still maintain the initial assumption that is specific to this model. That is, $\overline{\pi} = 2\pi$ and $\underline{\pi} = 0$ in view of the profit structure and π is large enough compared to β and u that $\frac{\pi}{\beta u} > 4$.

Supplementary note

We have assumed that all technological development schemes are both available to firms and implementable for a government authority. But as alternative cases, let us suppose that: (i) the government authority is unable to specify an optimal distribution in grant-back contracts; (ii) a grant-back clause and a RJV are unavailable and unimplementable due to institutional and legal inadequacy or costly arrangements. [45] With regard to (i) ([ii]), when consumer surplus is so negligible that $C < C_6 = \frac{\pi(2\sqrt{\beta u \pi} - \beta u)}{2\beta u}$ ($C < C_4 = \frac{\pi\sqrt{\pi}}{2\sqrt{\beta u}}$), a grant-back contract with a Nash bargaining solution (an appropriation without technology transfer), which saves resources allocated to the follow-on technology, is more socially desirable than a RJV (a license contract without a grant-back clause). Contrastingly, as consumer surplus is large enough that $C > C_6 = \frac{\pi(2\sqrt{\beta u \pi} - \beta u)}{2\beta u}$ ($C > C_4 = \frac{\pi\sqrt{\pi}}{2\sqrt{\beta u}}$), technology competition in follow-on innovation has a positive effect on social welfare, increasing the probability of improved final products appearing in the market. See Figure 1.3 that illustrates the highest social welfare along with the magnitude of consumer surplus.

1.5. Extension: uncertainty in initial innovation

An uncertainty factor has been so far included only in the follow-on innovation stage based on the assumption that the accomplishment of the follow-on technology is quite more difficult. But this assumption seems somewhat specific and restrictive while facilitating the analysis accordingly. Hence, a model is also investigated that accom-

[45] When $1 < \frac{\pi}{\beta u} < 4$ holds, firms cannot apply a grant-back clause due to relatively high costs and large uncertainty of developing technology D.

modates an uncertainty factor in the initial innovation stage, although the model is not only more complicated but also less conducive to providing a clear-cut result.

Let $m > 0$ denote the proportion of uncertainty included in initial innovation compared to follow-on innovation, so that mu represents the degree of uncertainty attached to developing technology R. That is, the larger the uncertainty factor, m, the more uncertain developing technology R compared to technology D. Other parameter settings and conditions are assumed to be still the same as before.

In the first place, we focus on the profit of a RJV maintaining the assumption that the two firms cannot extract any consumer surplus. Let us represent the third-stage profit of firm i (for $i = 1, 2$) as:

$$V_i^{J^*} = q\left[\frac{(2D_i^{J^*} + u)\pi}{D_1^{J^*} + D_2^{J^*} + u} - \beta D_i^{J^*}\right] + (1-q)\pi - \alpha R_i^J$$
$$\text{where } q = \frac{\sum_{n=1}^{2} R_n^J}{\sum_{n=1}^{2} R_n^J + mu}. \tag{1.41}$$

Now q equals the probability of the RJV achieving technology R in Stage 1. If the firms consolidating as a RJV fail to innovate technology R, they are unable to proceed to the next innovation stage of developing technology D, and as a consequence, earn the revenue of only π. The first-stage expected profit in Stage 1 earned from the formation of the RJV is defined as:

$$\Omega^J = \sum_{n=1}^{2} V_n^{J^*} = 2\left[q\left[\frac{(2D_i^{J^*} + u)\pi}{D_1^{J^*} + D_2^{J^*} + u} - \beta D_i^{J^*}\right] + (1-q)\pi\right] - \alpha R^J,$$
$$\text{where } R^J = \sum_{n=1}^{2} R_n^J \text{ and } q = \frac{R^J}{R^J + mu}. \tag{1.42}$$

Based on backward induction, the first-order condition of firm 1 with regard to D_1^J in Stage 3 is the same in accordance with the previous analysis: $D_1^{J^*} = D_2^{J^*} = \frac{\pi - \beta u}{2\beta}$. Substituting it into Equation (1.42), we can modify the equation such that:

$$\Omega^J = 2\left[q\left(\frac{\pi + \beta u}{2}\right) + (1-q)\pi\right] - \alpha R^J. \tag{1.43}$$

Because the profits of the two firms earned from the RJV are equivalent, the RJV will find optimal investment in technology R to maximize its profit.

Likewise the previous analysis where no uncertainty factor is included in the initial innovation stage, the RJV scheme inherently saves research resources channeled into the initial technology. And yet, the implication for cumulative innovation is totally different. In the previous cases, on the one hand, since investment in the initial technology is an entire waste of research resources due to the absence of uncertainty, firms can reduce it to the utmost limit in which the initial innovation is barely realized (i.e. infinitesimally small amount of investment, ε, in the analysis). On the other hand, the present case claims that the RJV cannot fundamentally conduct substantial investment in the initial technology due to uncertainty accruing to it. More precisely, the RJV optimally conducts investment of ε that maximizes the payoff in the present case. [46] However, with regard to the success probability of initial innovation, we derive $q = \frac{R^J}{R^J + mu} \to 0$ as $R^J = \varepsilon \to 0$. This suggests that as investment in the initial technology is infinitesimally small, initial innovation is unlikely to be successful. In turn, the follow-on innovation is not achieved either, which implies that cumulative

[46] As demonstrated in Section 1.7, the first derivative of Equation (1.43) with regard to R provides $\frac{\Omega^J}{R^J} < 0$ for any R^J. Hence, optimal R^J for the RJV scheme should be an infinitesimal value of ε.

innovation never occurs.

For this reason, when consumer surplus is significantly positive, the RJV may not be socially desirable in terms of social welfare, unless a government authority can force it to conduct significantly positive investment in the initial technology. This is because the success probability of initial innovation, q, would be rather low. In other words, the RJV scheme may deteriorate social welfare by nullifying potential consumer surplus earlier at the initial innovation stage. [47]

Lemma 1.3 If there exists uncertainty in initial innovation in addition to follow-on innovation, the RJV scheme is unlikely to achieve initial innovation due to the absence of sufficient investment in technology R. Hence, the improved final products as a result of cumulative innovation are unlikely to appear in the market in this technological development scheme.

In a similar fashion, the profit in each technological development scheme can be formulated. (Their detailed formulations are described in Section 1.7.) Proposition 1.6 finds optimal investment in technology R that maximizes profits and compares the amounts of investment, excluding investment provided by a grant-back contract with an optimal distribution, R^{O^*}, which was formulated in Subsection 1.3.4. Although the ranking of the amounts is retained, there are some points added that we should retain.

Proposition 1.6 With regard to investment in technology R:

(1) The amounts of investment are such that $R^{A^*} > R^{F^*} > R^{G^*} > R^{L^*} > R^{J^*} = \varepsilon$, and

[47] In reference to Footnote 23, when a RJV does not invest at all in technology R, social welfare in this scheme approximately equals 2π, where consumer surplus assumed to be zero. However, where there is large consumer surplus, potential social welfare loss resulting from a RJV could be large.

the “resilience to uncertainty" is strong in accordance with this ranking; and

(2) If an uncertainty factor included in the initial innovation stage is large enough, investment in technology R may not occur, and therefore, cumulative innovation may not be achieved.

When an uncertainty factor is included only in the follow-on innovation stage, investment in the initial technology does not provide any benefits, but only wastes research resources. But if we make an assumption that uncertainty also exists in initial innovation, implications are likely to vary. More precisely, the fiercer the initial competition, the higher the probability of firms achieving the initial technology. Remember that the initial technology is a decisive key to further innovations in later stages. In this way, technology competition for initial innovation prepares the ground for a series of cumulative innovation, whereby it bestows the potential to serve improved final products that are expected to increase consumer surplus.

As is shown by this proposition, an appropriation without technology transfer results in the highest probability of either firm achieving technology R, and thus, is the most “resilient to uncertainty" in initial innovation. (Recall that the probability is defined as $q = \frac{R}{R+mu}$ with $R = \sum_{n=1}^{2} R_n$, which is an increasing function in R.) In this light, it marks a favorable feature especially from the viewpoint of whether initial innovation succeeds. On the other hand, a RJV has the smallest incentive to innovate the initial technology and is the most “vulnerable to uncertainty" as we have already seen in Lemma 1.3. (It is demonstrated in Lemma 1.3 that the probability of achieving the initial technology is almost zero.)

We can derive a policy implication that is the exact opposite to the previous sections where uncertainty exists solely in follow-on innovation: if a government authority is to place much emphasis on the success in achieving the initial technology, it is feasible to allow firms to use the scheme of an appropriation without technology transfer. But it should be noted that since profits of the firms are also dependent on parameters such as uncertainty included in follow-on innovation, the firms may not choose the scheme that is most conducive to achieving initial innovation. Consequently, although it is necessary to at least create an environment in which the freedom to choose an appropriation is readily available, enforcement of the scheme by a government authority should still pose a challenge.

In addition, a license contract with a grant-back clause provides different implications. Among other things, a grant-back contract with a license fee generates a stronger incentive than other grant-back schemes, so that the former could be preferred in order to achieve the initial technology. This result is also totally opposite to that discussed in the previous sections. Furthermore, we can derive a similar intuition for investment in the initial technology, as we discussed the optimal investment in Proposition 1.3; when it is possible to specify a particular distribution of profits obtained from a grant-back contract, we may be able to successfully realize an initial innovation by appropriately controlling the incentive to develop the initial technology, to the extent that uncertainty accruing to developing the initial technology is not highly large. [48]

[48] This argument can be applied to the case of a license contract without a grant-back clause. (See Subsection 1.3.3.) When it is possible to specify a particular license fee, we may be able to successfully realize an initial innovation by appropriately controlling the incentive to develop the initial technology.

But alarmingly, since social welfare generally depends on various factors such as technological development costs, the degree of uncertainty that exists in both innovation stages, and the magnitude of consumer surplus, the decision of choosing optimal technological development schemes is much more complicated than it seems.

1.6. Concluding remarks

This chapter attempted to investigate which scheme of technological development is desirable in the stream of cumulative innovation mainly with an eye on the effects of a grant-back clause. Two firms compete for initial and follow-on technological development, where uncertainty is at first assumed to exist on the part of follow-on innovation. It was demonstrated that when consumer surplus obtained from cumulative innovation is negligible, there is a trade-off between investment in the initial and follow-on technologies according to technological development schemes. That is, if investment in the follow-on technology is conducted at the socially optimal level, overinvestment in the initial technology never fails to become exacerbated.

The study found that a grant-back contract combined with an appropriate distribution of expected profits (Nash bargaining solution; among other things, optimal grant-back contract) mitigates social overinvestment in the both initial and follow-on technologies, and thereby, improves social welfare. This result is led by the fact that such grant-back contracts not only decrease the overincentive for follow-on innovation by ensuring a licensor access to the follow-on technology, but also reduce the overincentive for the initial technology by appropriately controlling the attractiveness of winning initial technology competition, so that the common pool externalities in both innovation stages are mitigated. Among other things, it was shown that if a government authority can specify a particular distribution of profits between these two firms, socially optimal investment in the initial technology can be realized.

Furthermore, assuming significantly positive consumer surplus instead, it was revealed that competition in the follow-on technology creates higher social welfare, especially when consumer surplus is large. This implies that, by increasing the probability that improved final products enabled by the cumulative innovation appear in the market, the positive competition effect may overcome the overincentive problem caused by common pool externalities.

Finally, the model was extended so as to include uncertainty on the part of initial innovation in addition to follow-on innovation. The intuition was thereby derived that competition in the initial technology would be much more important for the advent of improved final products through building a basis for cumulative innovation.

The result that a grant-back clause plays a role in internalizing the common pool externalities (i.e. overcompetition) in the follow-on technology is similar to that derived by van Dijk (2000). The underlying mechanism is that since a licensor can require a licensee to grant back improvements achieved through the use of the licensed technology, both a licensor and a licensee weaken incentives to win competition in this technology. Nevertheless, there is a significant difference between these two studies; although van Dijk (2000) omits the effect of a grant-back clause on the incentive to innovate the initial technology, this study has explicitly incorporated it into the model. Moreover, this chapter has aimed to extend the argument to cases where consumer surplus being independent of firms' profits is significantly positive and where uncertainty is included in both initial and follow-on innovation.

Generally, intellectual property and antitrust law anticipate that license contracts contribute to disseminating licensors' state-of-the-art technologies to others. But it seems often underrated that some license contracts, especially a grant-back contract with a license fee, may greatly increase incentives to innovate the initial technology by making it more attractive to a licensor. When uncertainty does not exist or is relatively small in developing the initial technology, it may worsen the overincentive of firms to innovate first. In this viewpoint, a grant-back contract with a Nash bargaining solution or an optimal grant-back contract, which attaches a more distribution of the profits to a licensee, can be far more desirable than a grant-back contract with a license fee. On the other hand, we need to note that some debate remains left as to whether these desirable distributions are autonomously achievable without the support of a government authority.

In relation to a grant-back clause, this chapter highlights an important point; it increases or reduces the incentive for initial innovation by changing the attractiveness of the position becoming a licensor. Hence, if uncertainty is included in initial innovation, increased investment in the initial technology induced by technology competition through the control of the bargaining position may be justified, on ground that it can sow a seed of achieving cumulative innovation. This is why policymakers need to encourage firms to deliberately employ an appropriate technological development scheme, taking into account various factors such as the costs of technological development, the degree of uncertainty, and the magnitude of consumer surplus, all of which are imperative to the consequences of cumulative innovation.

What follows should be discussed as future challenges. The first challenge is a basic model setting. The Poisson arrival rate can be used as a success probability of innovations instead of rank-order tournament assumption and the decision of technological development schemes can be assumed endogenous in the model. Second, although we assume that firms can never extract consumer surplus, this assumption may be perceived as being restrictive. Another extreme case is perfect extraction, where the objective of a single firm to maximize profits and social welfare coincides. It would be therefore critical to examine how their incentives can be more closely connected to social welfare when extraction of consumer surplus is partly possible. Third, and more importantly, optimal investment in the initial technology was derived in the case where an uncertainty factor is included in the initial innovation stage, too. But this chapter has not conducted a full analysis of social welfare mainly due to the extreme complexity of the analytical model. A further effort should be made to establish a better improved cumulative innovation model as a way to evaluate which technological development scheme, and under which condition, is desirable.

1.7. Appendices

The mathematical demonstrations are gathered in this section. The proofs of Propositions and Lemmas are as follows.

Lemma 1.1 (1) $R^{A^*} - R^{L^*} = \frac{2\pi+\beta u-3\sqrt{\beta u\pi}}{2\alpha} - \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha} = \frac{\sqrt{\pi}(\sqrt{\pi}-\sqrt{\beta u})}{2\alpha} > 0 \Leftrightarrow R^{A^*} > R^{L^*}$ under $\frac{\pi}{\beta u} > 1$.

(2) $D^{L^*} - D^{A^*} = \frac{\pi-\beta u}{\beta} - (\sqrt{\frac{u\pi}{\beta}} - u) = \frac{\sqrt{\pi}(\sqrt{\pi}-\sqrt{\beta u})}{\beta} > 0 \Leftrightarrow D^{L^*} > D^{A^*}$ under $\frac{\pi}{\beta u} > 1$. ■

Proposition 1.1 We compare R^{G^*} with R^{A^*} and R^{L^*} for $\frac{\pi}{\beta u} > 4$. $R^{A^*} - R^{G^*} = \frac{2\pi+\beta u-3\sqrt{\beta u\pi}}{2\alpha} - \frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha} = \frac{\pi-2\sqrt{\beta u\pi}}{4\alpha} = \frac{\sqrt{\pi}(\sqrt{\pi}-2\sqrt{\beta u})}{4\alpha} > 0 \Leftrightarrow R^{A^*} > R^{G^*}$. In addition, $R^{G^*} - R^{L^*} = \frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{4\alpha} - \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{2\alpha} = \frac{\sqrt{\pi}(\sqrt{\pi}-2\sqrt{\beta u})}{4\alpha} > 0 \Leftrightarrow R^{G^*} > R^{L^*}$. We can therefore conclude that $R^{A^*} > R^{G^*} > R^{L^*} > R^{J^*}$. Next, we compare D^{G^*} with D^{A^*} and $D^{L^*}(= D^{J^*})$. $D^{G^*} - D^{A^*} = \frac{\pi-2\beta u}{2\beta} - (\sqrt{\frac{u\pi}{\beta}} - u) = \frac{\sqrt{\pi}(\sqrt{\pi}-2\sqrt{\beta u})}{2\beta} > 0 \Leftrightarrow D^{G^*} > D^{A^*}$. In addition, $D^{L^*} - D^{G^*} = \frac{\pi-\beta u}{\beta} - \frac{\pi-2\beta u}{2\beta} = \frac{\pi}{2\beta} > 0 \Leftrightarrow D^{L^*} > D^{G^*}$. These results lead to $D^{L^*} = D^{J^*} > D^{G^*} > D^{A^*}$. ■

Proposition 1.3 (1) By solving $R^{O^*}(k^*) = 0$ with regard to k, we obtain $k^* = \frac{-\pi-2\beta u+4\sqrt{\beta u\pi}}{2\pi}$. Next, consider the equation, $-\pi - 2\beta u + 4\sqrt{\beta u\pi} = 0$, which can be transformed into $f(\sqrt{\frac{\pi}{\beta u}}) = -(\sqrt{\frac{\pi}{\beta u}})^2 + 4(\sqrt{\frac{\pi}{\beta u}}) - 2 = 0$ by dividing the both sides of the equation by βu. Solving this quadratic equation provides $\sqrt{\frac{\pi}{\beta u}} = 2 \pm \sqrt{2}$, namely $\frac{\pi}{\beta u} = 6 + 4\sqrt{2}$ ($\simeq$ 11.657) or $6 - 4\sqrt{2}$ ($\simeq$ 0.343). Hence, $f(\sqrt{\frac{\pi}{\beta u}}) > 0$ holds for $6 - 4\sqrt{2} < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$. However, since $\frac{\pi}{\beta u} > 4$ is assumed in the grant-back case, $f(\sqrt{\frac{\pi}{\beta u}}) > 0$ holds only for $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$. Assuming $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$, we find a unique $k^* = \frac{-\pi-2\beta u+4\sqrt{\beta u\pi}}{2\pi} > 0$ that induces $R^{O^*}(k^*) = 0$ because $R^{O^*}(0) < 0$ and

$R^{O^*}(1) > 0$ (i.e. the intermediate-value theorem). If an approximate specification is given as $k^{O^*} \approx k^*$, optimal investment in technology R, such as $R^{J^*} = \varepsilon > 0$, can be achieved.

(2) We can demonstrate $k^* < \frac{1}{2}$ since $\frac{1}{2} - k^* = \frac{\pi+\beta u-2\sqrt{\beta u\pi}}{\pi} = \frac{(\sqrt{\pi}-\sqrt{\beta u})^2}{\pi} > 0$. ■

Proposition 1.4 (1) When $4 < \frac{\pi}{\beta u} < 6 + 4\sqrt{2}$, social welfare in each case can be derived as follows: $\Omega^{J^*} \simeq \pi + \beta u$; $\Omega^{A^*} = \frac{2\pi+\beta u+\sqrt{\beta u\pi}}{2}$; $\Omega^{L^*} = \frac{\pi+\beta u+2\sqrt{\beta u\pi}}{2}$; $\Omega^{G^*} = \frac{3\pi+2\beta u+4\sqrt{\beta u\pi}}{4}$; $\Omega^{N^*} = \frac{2\pi+\beta u+2\sqrt{\beta u\pi}}{2}$; and $\Omega^{O^*} \simeq \frac{3\pi+2\beta u}{2}$. Then, we obtain: $\Omega^{O^*} - \Omega^{N^*} = \frac{(\sqrt{\pi}-\sqrt{\beta u})^2}{2} > 0 \Leftrightarrow \Omega^{O^*} > \Omega^{N^*}$; $\Omega^{N^*} - \Omega^{A^*} = \frac{\sqrt{\beta u\pi}}{2} > 0 \Leftrightarrow \Omega^{N^*} > \Omega^{A^*}$; $\Omega^{A^*} - \Omega^{G^*} = \frac{\sqrt{\pi}(\sqrt{\pi}-2\sqrt{\beta u})}{4} > 0 \Leftrightarrow \Omega^{A^*} > \Omega^{G^*}$; $\Omega^{G^*} - \Omega^{J^*} = \frac{-\pi-2\beta u+4\sqrt{\beta u\pi}}{4} > 0 \Leftrightarrow \Omega^{G^*} > \Omega^{J^*}$; and $\Omega^{J^*} - \Omega^{L^*} = \frac{(\sqrt{\pi}-\sqrt{\beta u})^2}{2} > 0 \Leftrightarrow \Omega^{J^*} > \Omega^{L^*}$ under the condition. Hence, $\Omega^{O^*} > \Omega^{N^*} > \Omega^{A^*} > \Omega^{G^*} > \Omega^{J^*} > \Omega^{L^*}$ is demonstrated.

(2) When $\frac{\pi}{\beta u} > 6 + 4\sqrt{2}$, social welfare of an optimal grant-back contract changes into $\Omega^{O^*} = \frac{5\pi+2\beta u+4\sqrt{\beta u\pi}}{4}$. Comparing Ω^{O^*} and Ω^{N^*} results in $\Omega^{O^*} - \Omega^{N^*} = \frac{\pi}{4} > 0 \Leftrightarrow \Omega^{O^*} > \Omega^{N^*}$. In addition, $\Omega^{A^*} - \Omega^{J^*} = \frac{\sqrt{\beta u}(\sqrt{\pi}-\sqrt{\beta u})}{2} > 0 \Leftrightarrow \Omega^{A^*} > \Omega^{J^*}$; $\Omega^{J^*} - \Omega^{G^*} = \frac{\pi+2\beta u-4\sqrt{\beta u\pi}}{4} > 0 \Leftrightarrow \Omega^{J^*} > \Omega^{G^*}$; and $\Omega^{G^*} - \Omega^{L^*} = \frac{\pi}{4} > 0 \Leftrightarrow \Omega^{G^*} > \Omega^{L^*}$. Hence, $\Omega^{O^*} > \Omega^{N^*} > \Omega^{A^*} > \Omega^{J^*} > \Omega^{G^*} > \Omega^{L^*}$ is demonstrated. ■

Proposition 1.5 According to the definition of Equation (1.40), social welfare in each case for $4 < \frac{\pi}{\beta u} < 6+4\sqrt{2}$ is formulated in what follows: $W^{J^*} \simeq \pi+\beta u+(1-\frac{\beta u}{\pi})C$; $W^{A^*} = \frac{2\pi+\beta u+\sqrt{\beta u\pi}}{2} + (1 - \sqrt{\frac{\beta u}{\pi}})C$; $W^{L^*} = \frac{\pi+\beta u+2\sqrt{\beta u\pi}}{2} + (1 - \frac{\beta u}{\pi})C$; $W^{G^*} = \frac{3\pi+2\beta u+4\sqrt{\beta u\pi}}{4} + (1 - \frac{2\beta u}{\pi})C$; $W^{N^*} = \frac{2\pi+\beta u+2\sqrt{\beta u\pi}}{2} + (1 - \frac{2\beta u}{\pi})C$; and $W^{O^*} = \frac{3\pi+2\beta u}{2} + (1 - \frac{2\beta u}{\pi})C$. By solving the simultaneous equations, we can obtain the intersections of these lines, from C_1 to

C_8 depicted in Figure 1.3. ■

Lemma 1.2 From Equation (1.43), the first-stage profit of the RJV is determined by:

$$\begin{aligned}\Omega^J &= q\left(\frac{\pi + \beta u}{2}\right) + (1-q)\pi - \alpha R^J \\ &= 2\left[\frac{R^J}{R^J + mu}\left(\frac{\pi + \beta u}{2}\right) + \frac{mu\pi}{R^J + mu}\right] - \alpha R^J. \end{aligned} \tag{1.44}$$

We examine the first derivative of Ω^J with regard to R^J:

$$\frac{\partial \Omega^J}{\partial R^J} = 2\left[\frac{mu}{(R^J + mu)^2}\left(\frac{\pi + \beta u}{2}\right) - \frac{mu\pi}{(R^J + mu)^2}\right] - \alpha = \frac{mu(\beta u - \pi)}{(R^J + mu)^2} - \alpha. \tag{1.45}$$

From Equation (1.45), we find that $\frac{\partial \Omega^J}{\partial R^J} < 0$ under $\frac{\pi}{\beta u} > 0$. Since it is assumed that the RJV conducts positive investment in technology R, optimal investment is derived as $R^{J^*} = \varepsilon$ that is an infinitesimally small value. However, the probability of achieving technology R converges to $q = \frac{R^J}{R^J + mu} \to 0$ as $R^J \to 0$. ■

Proposition 1.6 (1) The first-stage profits are demonstrated in what follows. In the case of an appropriation without technology transfer:

$$\begin{aligned}\Omega_1^A &= \frac{R_1^A}{R_1^A + R_2^A + mu}\left[\frac{(2D_1^{A^A} + u)\pi}{D_1^{A^*} + u} - \beta D_1^{A^*}\right] + \frac{R_2^A}{R_1^A + R_2^A + mu}\left[\frac{u\pi}{D_1^{A^*} + u}\right] \\ &\quad + \frac{mu\pi}{R_1^A + R_2^A + mu} - \alpha R_1^A \\ &= \frac{R_1^A(2\pi + \beta u - 2\sqrt{\beta u \pi})}{R_1^A + R_2^A + mu} + \frac{R_2^A(\sqrt{\beta u \pi})}{R_1^A + R_2^A + mu} + \frac{mu\pi}{R_1^A + R_2^A + mu} - \alpha R_1^A. \end{aligned} \tag{1.46}$$

The first-order condition is:

$$\frac{\partial \Omega_1^A}{\partial R_1^A} = \frac{R_2^A(2\pi + \beta u - 3\sqrt{\beta u \pi})}{(R_1^A + R_2^A + mu)^2} + \frac{mu(\pi + \beta u - 2\sqrt{\beta u \pi})}{(R_1^A + R_2^A + mu)^2} - \alpha = 0. \tag{1.47}$$

From the symmetry, $R_1^{A^*} = R_2^{A^*}$, we obtain optimal investment such that:

$$\begin{aligned} R_1^{A^*} = R_2^{A^*} &= \frac{2\pi + \beta u - 3\sqrt{\beta u \pi} - 4\alpha mu}{8\alpha} \\ &+ \frac{\sqrt{(2\pi + \beta u - 3\sqrt{\beta u \pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - (\sqrt{\pi} - \sqrt{\beta u})^2]}}{8\alpha} \\ &\equiv FT^{A^*} + ST^{A^*}. \end{aligned} \tag{1.48}$$

Taking into account the second-order condition, the larger root can be selected as a solution. (The same applies hereafter.)

In the case of a license contract without a grant-back clause:

$$\begin{aligned} \Omega_1^L &= \frac{R_1^L}{R_1^L + R_2^L + mu}\left[\frac{(2D_1^{L^*} + u)\pi}{D_1^{L^*} + D_2^{L^*} + u} - \beta D_1^{L^*} + f^{L^*}\right] \\ &+ \frac{R_2^L}{R_1^L + R_2^L + mu}\left[\frac{(2\tilde{D}_1^{L^*} + u)\pi}{\tilde{D}_1^{L^*} + \tilde{D}_2^{L^*} + u} - \beta \tilde{D}_1^{L^*} - f^{L^*}\right] + \frac{mu\pi}{R_1^L + R_2^L + mu} - \alpha R_1^L \\ &= \frac{R_1^L(\pi + \beta u - \sqrt{\beta u \pi})}{R_1^L + R_2^L + mu} + \frac{R_2^L(\sqrt{\beta u \pi})}{R_1^L + R_2^L + mu} + \frac{mu\pi}{R_1^L + R_2^L + mu} - \alpha R_1^L. \end{aligned} \tag{1.49}$$

The first-order condition is:

$$\frac{\partial \Omega_1^L}{\partial R_1^L} = \frac{R_2^L(\pi + \beta u - 2\sqrt{\beta u \pi})}{(R_1^L + R_2^L + mu)^2} + \frac{mu(\beta u - \sqrt{\beta u \pi})}{(R_1^L + R_2^L + mu)^2} - \alpha = 0. \tag{1.50}$$

We obtain optimal investment such that:

$$\begin{aligned} R_1^{L^*} = R_2^{L^*} &= \frac{\pi + \beta u - 2\sqrt{\beta u \pi} - 4\alpha mu}{8\alpha} \\ &+ \frac{\sqrt{(\pi + \beta u - 2\sqrt{\beta u \pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - (\beta u - \sqrt{\beta u \pi})]}}{8\alpha} \\ &\equiv FT^{L^*} + ST^{L^*}. \end{aligned} \tag{1.51}$$

In the case of a grant-back contract with a license fee:

$$\Omega_1^G = \frac{R_1^G}{R_1^G + R_2^G + mu}\left[\frac{(2D_1^{G^*} + u)\pi}{D_1^{G^*} + D_2^{G^*} + u} - \beta D_1^{G^*}\right] + \frac{R_2^G}{R_1^G + R_2^G + mu}\left[\frac{(2\tilde{D}_1^{G^*} + u)\pi}{\tilde{D}_1^{G^*} + \tilde{D}_2^{G^*} + u} - \beta \tilde{D}_1^{G^*}\right] + \frac{mu\pi}{R_1^G + R_2^G + mu} - \alpha R_1^G$$
$$= \frac{R_1^G}{R_1^G + R_2^G + mu}\left(\frac{3\pi + 2\beta u - 2\sqrt{\beta u\pi}}{2}\right) + \frac{R_2^G(\sqrt{\beta u\pi})}{R_1^G + R_2^G + mu} + \frac{mu\pi}{R_1^G + R_2^G + mu} - \alpha R_1^G. \tag{1.52}$$

The first order condition is:

$$\frac{\partial \Omega_1^G}{\partial R_1^G} = \frac{R_2^G}{(R_1^G + R_2^G + mu)^2}\left(\frac{3\pi + 2\beta u - 4\sqrt{\beta u\pi}}{2}\right) + \frac{mu}{(R_1^G + R_2^G + mu)^2}\left(\frac{\pi + 2\beta u - 2\sqrt{\beta u\pi}}{2}\right) - \alpha = 0. \tag{1.53}$$

We obtain optimal investment such that:

$$R_1^{G^*} = R_2^{G^*} = \frac{\frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{2} - 4\alpha mu}{8\alpha} + \frac{\sqrt{(\frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{2} - 4\alpha mu)^2 - 16\alpha mu(\alpha mu - \frac{\pi+2\beta u-2\sqrt{\beta u\pi}}{2})}}{8\alpha} \equiv FT^{G^*} + ST^{G^*}. \tag{1.54}$$

Finally, in the case of a grant-back contract with the Nash bargaining solution:

$$\Omega_1^N = \frac{R_1^N}{R_1^N + R_2^N + mu}\left(\frac{5\pi}{4} + \beta u - \sqrt{\beta u\pi}\right) + \frac{R_2^N}{R_1^N + R_2^N + mu}\left(\frac{\pi}{4} + \sqrt{\beta u\pi}\right) + \frac{mu\pi}{R_1^N + R_2^N + mu} - \alpha R_1^N. \tag{1.55}$$

The first-order condition is:

$$\frac{\partial \Omega_1^N}{\partial R_1^N} = \frac{R_2^N(\pi + \beta u - 2\sqrt{\beta u \pi})}{(R_1^N + R_2^N + mu)^2} + \frac{mu}{(R_1^N + R_2^N + mu)^2}\left(\frac{\pi}{4} + \beta u - \sqrt{\beta u \pi}\right) - \alpha = 0. \quad (1.56)$$

We obtain optimal investment such that:

$$\begin{aligned} R_1^{N^*} = R_2^{N^*} &= \frac{\pi + \beta u - 2\sqrt{\beta u \pi} - 4\alpha mu}{8\alpha} \\ &+ \frac{\sqrt{(\pi + \beta u - 2\sqrt{\beta u \pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - \frac{(\sqrt{\pi} - 2\sqrt{\beta u})^2}{4}]}}{8\alpha} \\ &\equiv FT^{N^*} + ST^{N^*}. \end{aligned} \quad (1.57)$$

The amount of each investment by firm 1 is compared without loss of generality:

$$\begin{aligned} R_1^{A^*} - R_1^{G^*} &= (FT^{A^*} - FT^{G^*}) + (ST^{A^*} - ST^{G^*}) \\ &= \frac{\pi - 2\sqrt{\beta u \pi}}{16\alpha} \\ &+ \frac{\sqrt{(2\pi + \beta u - 3\sqrt{\beta u \pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - (\sqrt{\pi} - \sqrt{\beta u})^2]}}{8\alpha} \\ &- \frac{\sqrt{(\frac{3\pi + 2\beta u - 4\sqrt{\beta u \pi}}{2} - 4\alpha mu)^2 - 16\alpha mu(\alpha mu - \frac{\pi + 2\beta u - 2\sqrt{\beta u \pi}}{2})}}{8\alpha}. \end{aligned} \quad (1.58)$$

Since it has been already demonstrated that $\pi - 2\sqrt{\beta u \pi} = \sqrt{\pi}(\sqrt{\pi} - 2\sqrt{\beta u}) > 0$, it is sufficient to check the following equation: $-16\alpha mu[\alpha mu - (\sqrt{\pi} - \sqrt{\beta u})^2] + 16\alpha mu(\alpha mu - \frac{\pi + 2\beta u - 2\sqrt{\beta u \pi}}{2}) = 8\alpha mu(\pi - 2\sqrt{\beta u \pi}) = 8\alpha mu\sqrt{\pi}(\sqrt{\pi} - 2\sqrt{\beta u}) > 0$ for $\frac{\pi}{\beta u} > 4$. Hence, we obtain $R_1^{A^*} > R_1^{G^*}$.

The same calculations are applied hereafter.

$$R_1^{G^*} - R_1^{N^*} = (FT^{G^*} - FT^{N^*}) + (ST^{G^*} - ST^{N^*})$$

$$= \frac{\pi}{16\alpha} + \frac{\sqrt{(\frac{3\pi+2\beta u-4\sqrt{\beta u\pi}}{2} - 4\alpha mu)^2 - 16\alpha mu(\alpha mu - \frac{\pi+2\beta u-2\sqrt{\beta u\pi}}{2})}}{8\alpha}$$
$$- \frac{\sqrt{(\pi + \beta u - 2\sqrt{\beta u\pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - \frac{(\sqrt{\pi}-2\sqrt{\beta u})^2}{4}]}}{8\alpha}. \quad (1.59)$$

Since $-16\alpha mu(\alpha mu - \frac{\pi+2\beta u-2\sqrt{\beta u\pi}}{2}) + 16\alpha mu[\alpha mu - \frac{(\sqrt{\pi}-2\sqrt{\beta u\pi})^2}{4}] = 4\alpha mu\pi > 0$, we obtain $R_1^{G^*} > R_1^{N^*}$.

$$R_1^{N^*} - R_1^{L^*} = (FT^{N^*} - FT^{L^*}) + (ST^{N^*} - ST^{L^*})$$
$$= \frac{\sqrt{(\pi + \beta u - 2\sqrt{\beta u\pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - \frac{(\sqrt{\pi}-2\sqrt{\beta u})^2}{4}]}}{8\alpha}$$
$$- \frac{\sqrt{(\pi + \beta u - 2\sqrt{\beta u\pi} - 4\alpha mu)^2 - 16\alpha mu[\alpha mu - (\beta u - \sqrt{\beta u\pi})]}}{8\alpha}. \quad (1.60)$$

Since $-16\alpha mu[\alpha mu - \frac{(\sqrt{\pi}-2\sqrt{\beta u})^2}{4}] + 16\alpha mu[\alpha mu - (\beta u - \sqrt{\beta u\pi})] = 16\alpha mu(\sqrt{\beta u\pi} - \beta u) + 4\alpha mu\pi > 0$, we obtain $R_1^{N^*} > R_1^{L^*}$. In addition, $R^{L^*} > R^{J^*} = \varepsilon \approx 0$ generally holds. By consolidating the above, we can conclude $R^{A^*} > R^{G^*} > R^{N^*} > R^{L^*} > R^{J^*} \approx 0$.

(2) We can show that R^{A^*}, R^{L^*}, R^{G^*}, and R^{N^*} are all decreasing in m. Hence, if m is sufficiently large, these values cannot take neither positive values nor real numbers. ■

References

[1] Aghion, P. and Howitt, P. (1992), “A Model of Growth through Creative Destruction", *Econometrica*, 60(2), pp. 323-351

[2] Arrow, K. J. (1962), “Economics Welfare and the Allocation of Resources for Innovation", Chapter 23 in *The Rate and Direction of Inventive Activity*, ed. Nelson, R. R., Princeton University Press, Princeton, NJ, pp. 609-626

[3] Banal-Estañol, A. and Macho-Stadler, I. (2010), “Scientific and Commercial Incentives in R&D: Research versus Development?", *Journal of Economics and*

Management Strategy, 19(1), pp. 185-221

[4] Barzel, Y. (1968), “Optimal Timing of Innovations", *Review of Economics and Statistics*, 50(3), pp. 348-355

[5] Bessen, J., Ford, J. L., and Meurer, M. J. (2011), “The Private and Social Costs of Patent Trolls", *Regulation*, 34(4), pp. 26-35

[6] Bloom, N. (2014), “Fluctuations in Uncertainty", *Journal of Economic Perspectives*, 28(2), pp. 153-176

[7] Choi, J. P. (2002), “A Dynamic Analysis of Licensing: The ‘Boomerang’ Effect and Grant- Back Clauses", *International Economic Review*, 43(3), pp. 803-829

[8] Coase, R. H. (1960), “The Problem of Social Cost", *Journal of Law and Economics*, 3(1), pp. 1-44

[9] Cohen, L., Gurun, U., and Kominers, S. D. (2014), “Patent Trolls: Evidence from Targeted Firms", NBER Working Paper Series, Paper No. 20322, National Bureau of Economic Research, available at: http://www.nber.org/papers/w20322.pdf

[10] Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000), “Protecting their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)", NBER Working Paper Series, Paper No. 7552, National Bureau of Economic Research, available at: http://www.nber.org/papers/w7552.pdf

[11] Denicolò, V. (2000), “Two-Stage Patent Races and Patent Policy", *Rand Journal of Economics*, 31(3), pp. 488-501

[12] European Union (2004), “Commission Note: Guidelines on the Application of Article 81 of the RC Treaty to Technology Transfer Agreements", *Official Journal of the European Union*, 2004/C 101/02, available at: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52004XC0427(01)

[13] Green, J. R. and Scotchmer, S. (1995), “On the division of Profit in Sequential Innovation", *Rand Journal of Economics*, 26(1), pp. 20-33

[14] Grossman, G. M. and Helpman, E. (1991a), “Quality Ladders and Product Cycles", *Quarterly Journal of Economics*, 106(2), pp. 557-586

[15] Grossman, G. M. and Helpman, E. (1991b), “Quality Ladders and the Theory of Growth", *Review of Economic Studies*, 58(1), pp. 43-61

[16] Hall, B. (2007), “Patents and Patent Policy", *Oxford Review of Economic Policy*, 23(4), pp. 568-587

[17] Hatanaka, K. (2012), “An Economic Analysis of Grant-back Clauses" (in Japanese), CPRC Discussion Paper Series, Paper No. CPDP-56-J, Competition Policy Research Center, Japan Fair Trade Commission, available at: http://www.jftc.go.jp/cprc/discussionpapers/h24/cpdp_56_j_abstract.html

[18] Jalonen, H. (2012), “The Uncertainty of Innovation: A Systematic Review of the Literature", *Journal of Management Research*, 4(1), pp. 1-47

[19] Japan Fair Trade Commission (2016), “Guidelines on the Antitrust Laws regarding the Use of Intellectual Property" (in Japanese), available at: http://www.jftc.go.jp/dk/guideline/unyoukijun/chitekizaisan.html

[20] Johansen, H. (2012), “The Uncertainty of Innovation: A Systematic Review of the Literature", *Journal of Management Research*, 4(1), pp. 1-47

[21] Kamien, M. I., Muller, E., and Zang, I. (1992), “Research Joint Ventures and R&D Cartels", *American Economic Review*, 82(5), pp. 1293-1306

[22] Leone, M. I. and Reichstein, T. (2012), “Licensing-in Fosters Rapid Innovation! The Effect of the Grant-back Clause and Technological Unfamiliarity", *Strategic Management Journal*, 33(8), pp. 965-985

[23] Lévêque, F. and Ménière, Y. (2004), *The Economics of Patents and Copyright*, Berkeley Electronic Press, Berkeley, CA, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=642622

[24] Ministry of Economy, Trade and Industry of Japan (2011), “Survey of Efficiency in R&D of Japanese Firms: Quantitative Analysis of Open Innovation" (in Japanese), available at: http://www.meti.go.jp/meti_lib/report/2011fy/0022757.pdf

[25] Miyagawa, K. (2007), “Collusion and Research Joint Ventures", ISER Discussion Papers, Paper No. 704, Institute of Social and Economic Research, Osaka University, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1072464

[26] Moreira, S., Laursen, K. Leone, M. I., and Reichstein, T. (2012), “Cooperation or Competition: Grant-Back Clauses in Technology Licensing", paper presented at the DRUID Academy 2012, The Moeller Centre of University of Cambridge, January 2012, available at: http://druid8.sit.aau.dk/acc_papers/ain1su9m8x2tjii0rtlonyktvidl.pdf

[27] Nash, J. F. (1950), “The Bargaining Problem", *Econometrica*, 18(2), pp.155-162

[28] Nordhaus, W. (1969), *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change*, The MIT Press, Cambridge, MA

[29] O'Donoghue, T. (1998), "A Patentability Requirement for Sequential Innovation", *Rand Journal of Economics*, 29(4), pp. 654-679

[30] Osborne, M. J. and Rubinstein, A. (1990), *Bargaining and Markets*, Academic Press, San Diego, CA, pp. 69-90

[31] Reinganum, J. F. (1989), "The Timing of Innovation: Research, Development, and Diffusion", Chapter 14 in *Handbook of Industrial Organization*, Volume 1, eds. Schmalensee, R. and Willig, R. D., Elsevier, Amsterdam, pp. 849-908

[32] Rockett, K. (2010), "Property Rights and Innovation", Chapter 7 in *Handbook of the Economics of Innovation*, Volume 1, eds. Hall, B. and Rosenberg N., Elsevier, Amsterdam, pp. 315-380

[33] Romer, P. (1986), "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, 94(5), pp. 1002-1037

[34] Romer, P. (1990), "Endogenous Technological Change", *Journal of Political Economy*, 98(5), pp. 71-102

[35] Saur-Amaral, I. and Borges Gouveia, J. J. (2007), "University in Drug Development: Insights from a Portuguese Firm", *International Journal of Technology Intelligence and Planning*, 3(4), pp. 355-375

[36] Schumpeter, J. A. (1942), *Capitalism, Socialism, and Democracy*, Routledge, London

[37] Scotchmer, S. (1991), "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law", *Journal of Economic Perspectives*, 5(1), pp.29-41

[38] Scotchmer, S. (2004), *Innovation and Incentives*, The MIT Press, Cambridge, MA, pp. 127-159

[39] Shapiro, C. (1985), "Patent Licensing and R&D Rivalry", *American Economic Review*, 75(2), pp. 25-30

[40] Solow, R. M. (1957), "Technical Change and the Aggregate Production Function", *Review of Economics and Statistics*, 39(3), pp. 312-320

[41] Tao, Z. and Wu, C. (1997), "On the Organization of Cooperative Research and Development: Theory and Evidence", *International Journal of Industrial Organization*, 15(5), pp.573-596

[42] Tirole, J. (1989), *The Theory of Industrial Organization*, The MIT Press, Cambridge, MA, pp. 389-421

[43] van Dijk, T. (2000), "License Contracts, Future Exchange Clauses, and Technological Competition", *European Economic Review*, 44(8), pp. 1431-1448

Chapter 2. Grant-Backs and Attributes of Innovation: “But for . . . " Defense of a Grant-Back Clause

2.1. Introduction

Patent-holders are not always placed in an ideal position to exploit their own technologies in a market. In particular, while a patent-holder may be quite competitive in her own home market, she may not possess the required local expertise in order to perform well in other geographical markets. Alternatively, it could be that utilizing existing production technologies of other firms may be more efficient than serving other markets through export or direct investment in new technologies adjusted for different markets. For these reasons, licensing is a common way to transfer technologies, and thereby, to earn revenue from royalty payments. Licensing is also generally acknowledged as socially desirable because it disseminates innovations to other firms. Based upon a survey on licensing activities in the European Union (EU) and Japan conducted in the second half of 2007, Zuniga and Guellec (2009) observe not only that 20% of firms in Europe and 27% in Japan grant licenses to non-affiliated entities, but also that licensed firms tend to license in a large proportion of their patent portfolio with a figure above 80% or higher being common.

In addition to conventional license contracts, the focus of Chapter 2 is on a grant-back clause with a different perspective from Chapter 1. As previously mentioned in

Chapter 1, the grant-back clause provides a licensor with grant-back access to future innovation achieved by a licensee, and it has been in popular use. Cockburn (2007) reveals that 43% of license contracts of firms in his sample contain such grant-back clauses. Moreover, on the basis of econometrics research, Moreira et al. (2012) remark that grant-back clauses tend to be more commonly used within firms that are in the identical product market and that are familiar with relevant technologies.

In spite of the widespread use of grant-backs, their legal status has remained somewhat unclear. [1] In particular, it is worth noting the distinctive practice isolated in the previous so-called *Technology Transfer Guidelines* formulated by the EU in 2004 (the EU, 2004) that calls for certain types of grant-backs to come under careful scrutiny. OECD (2004) posits that some grant-back arrangements are more likely to damage a licensee's incentive for follow-on innovation and to cause more serious competitive problems than other licensing schemes. With this in mind, the EU (2004) explicitly distinguishes between “severable" and “non-severable" innovation. Whereas the definition of “severable" innovation is that the innovation can be employed without infringing the licensed technology, “non-severable" innovation is defined as such that the innovation cannot be exploited by a licensee in the absence of a licensor's permission. In line with this, the EU (2004) prescribes that from the viewpoint of the

[1] In the U.S., the Transwarp case established that the provision of grant-backs was not *per se* illegal. See Stokes & Smith Co. vs. Transparent-Wrap Machine Corp., 156 F.2d 198, 2d Cir. 1946 and Transparent-Wrap Machine Corp. vs. Stokes & Smith, 329 US 637, 1947. In subsequent cases, it was similarly judged that grant-backs were not necessarily inconsistent with competition policies unless they were part of a more general pattern of anticompetitive behaviors. See the United States vs. Switzer Bros., Inc, 1953 Trade Case, Para 67,598 (N.D. Cal. 1953) where the grant-back is one of several offensive items. Other examples include the United States vs. Imperial Chemical Industries, Ltd., 105 F. Supp. 215 (S.D.N.Y, 1952) and the United States vs. Besser Mfg., Co., 96 F. Supp. 304 (E.D. Mich. 1951).

attributes of innovation, grant-back clauses concerning severable innovation should be viewed from a more critical standpoint, especially when the grant-backs are exclusive. [2] Hence, the EU (2004) maintains that since grant-back clauses for severable innovation are more likely to severely impair the licensee's incentive to innovate further than for non-severable innovation, they should be regarded as socially undesirable. [3]

This chapter intends to raise a question about the above-mentioned critical views and to shed light on the counter-argument for including a grant-back clause, namely the "but for" defense. More precisely, if there is a concern held by an original licensor that a licensee may stop royalty payments, or in the worst possible scenario, may erode the licensor's profit by using the licensed technology to leap frog or to invent around the original patent, the licensor is unlikely to provide her original technology without a grant-back clause. The aim of this chapter is, therefore, to further investigate how the "but for" defense of a grant-back clause can be declared legitimate from the perspective of preserving a licensor's incentive (in a specific case, a licensee's incentive), when the analysis explicitly considers the attributes of innovation. Particularly, this study assesses whether the critical approach of the EU (2004) to grant-back clauses

[2] Article 109 of the EU (2004) says: "Article 5(1)(a) and 5(1)(b) concerns exclusive grant backs or assignments to the licensor of severable improvements of the licensed technology. ··· An obligation to grant the licensor an exclusive licence to severable improvements of the licensed technology or to assign such improvements to the licensor is likely to reduce the licensee's incentive to innovate since it hinders the licensee in exploiting his improvements, including by way of licensing to third parties. This is the case both where the severable improvement concerns the same application as the licensed technology and where the licensee develops new applications of the licensed technology".

[3] The current *Technology Transfer Guidelines* (the EU, 2014) is still somewhat characterized by skepticism toward grant-back clauses that are materialized in the form of exclusive grant-backs. (An exclusive grant-back is defined as a grant-back which prevents the licensee from exploiting the improvement either for its own production or for licensing out to third parties.) Article 129 of the EU (2014) says: "An obligation to grant the licensor an exclusive licence to improvements of the licensed technology or to assign such improvements to the licensor is likely to reduce the licensee's incentive to innovate since it hinders the licensee in exploiting the improvements, including by way of licensing to third parties".

for severable innovation can be justified.

The following three questions are addressed in this chapter. First, is a grant-back clause always indispensable in facilitating licensing activities? Second, does a grant-back clause affect a licensee's incentive to innovate? Third, is a grant-back likely to allow third parties to gain access to future innovations? Note that these questions at issue are totally different from those discussed in Chapter 1, which examines the manner in which grant-back clauses affect incentives in cumulative innovation induced by technology competition.

Let me briefly respond to these three questions. For the first question, it is demonstrated that in the case of a license contract with a single licensee, the “but for" argument is applied more firmly to severable innovation than to non-severable innovation. In other words, whereas a grant-back clause does not further facilitate any licensing activities for non-severable innovation, it does for severable innovation in some cases. The reason is briefly as follows: although severable innovation, which is not infringing the original patent, makes it difficult for a licensor to recapture the fruit of the follow-on innovation created by a licensee, a grant-back clause ensures the licensor's use of it without any explicit royalty payment.

With regard to the second question, grant-back clauses do not necessarily extinguish a licensee's motivation to innovate. This is because the licensor, who is generally assumed to have total bargaining power over including a grant-back clause, still can expect the licensee to complete the innovation if it is profitable to do so. Consequently, it is reasonable for the licensor to leave monetary motivation with the licensee in anticipation that she is willing to innovate further. Furthermore, this chapter also shows that when an innovation cost is strictly positive, the above-mentioned argument becomes even much more relevant.

Finally, in order to answer the third question, we need to add another potential licensee to the model. The situation becomes somewhat complicated in this instance. Considering multiple heterogeneous licensees in terms of innovation abilities (that is, the one is capable to innovate, and the other is not), we find that a licensor will be unable to license her original technology to both competent and incompetent licensees for severable innovation if territorial restrictions are in place. But grant-back clauses can open the door to licensing activities, whereby the competent licensee can innovate and the other incompetent licensee is certain to take advantage of the innovation.

The effect of a grant-back clause has been examined by some authors, as surveyed in Chapter 1, but they have not directed their attention to attributes of innovation. It is worth commenting that our modeling approach does not share some of the issues proposed by previous studies that have important implications for their conclusions. This chapter does not deal concretely with the problem of “overincentives" to innovate addressed by Chapter 1, which argues that grant-back contracts can lead to a socially desirable reduction in innovation incentives. In addition, contrary to the framework of Choi (2002), this study does not build on asymmetric information between a licensor and a licensee regarding the quality of licensed technologies. While Choi’s (2002) results illustrate that grant-back clauses help a licensor retain the incentive to transfer the state-of-the-art technology, the role of asymmetric information in this transfer is not the focus here. This is because, as described in the beginning of this section that mentions Zuniga and Guellec (2009), many grant-back licensing activities can be observed between firms that are familiar with relevant technologies to their businesses. Indeed, the issues analyzed in these previous studies can be applied to some cases, but the model of this chapter rules them out.

The main focus of this chapter is on the contrasting incentive for innovation according to the different attributes of innovation. By establishing the “but for" defense of a grant-back clause for severable innovation, we attempt to critically examine the EU (2004), which was more lenient toward non-severable innovation than severable innovation. To this end, it is proven that a licensor is likely to benefit from the grant-back clause as part of technology transfer especially for severable innovation. Thus, the contribution of this chapter to literatures on economics of intellectual property rights is that new light is shed on the characteristics of a grant-back clause, and thereby, the unique implications are driven for the effect of a grant-back on innovation incentives of a licensor as well as a licensee.

After this introduction, Section 2.2 outlines the model structure in a multistage game. Section 2.3 investigates the case of a single licensee in accordance with the attributes of innovation. Section 2.4 extends the basic model by changing its assumptions on territorial restriction and innovation cost. Section 2.5 analyzes the case of multiple heterogeneous licensees. Finally, in Section 2.6, the conclusion is presented followed with appendices of Section 2.7 and full references.

2.2. Model structure

In Section 2.2, a baseline model is established where one licensor and one licensee exist. Afterward in Section 2.4, the model will be extended to two heterogeneous licensees in terms of innovation abilities, with more varied implications.

It is assumed that a licensor (denoted by firm L) has already possessed a base technology (denoted by BT) that can be licensed to a licensee (denoted by firm A). For a descriptive purpose, the set of firms is defined as $F = \{L, A\}$. BT is assumed to be an original vital technology that leads to follow-on innovation. It is posited that only firm A is able to achieve innovation and develop the improved technology (denoted by IT) by utilizing BT. [4]

Whereas firms L and A sell products in their local markets without incurring any costs, they need to incur an additional cost, $c > 0$, to sell in another market. Thus, c can be regarded as a market entry cost, or a transportation cost that measures the distance between the two markets. It is reasonable to assume a limit on the range of c so that the firms can find it profitable to sell in all markets, which suggests that c must be a finite value. In addition, when an initial license contract is concluded, firm A innovates using BT at the cost of $k \geq 0$. It is postulated that whereas the willingness to pay by consumers in each market is measured as 1 with BT, it rises to $1 + \theta$ (with $\theta > 0$) if the products are supplied to consumers with IT. In this sense, IT adds on an

[4] As is frequently observed, inventing firms are not necessarily the best entities that can make full use of their inventions. Throughout history, follow-on inventors have succeeded in applying achievements of their predecessors to their interests, and thereby, have contributed to development of science and industrialization as a result. For example, steam engines newly invented by James Watt in the 19th century were utilized in various industrial fields such as steamboats and steam locomotives in generations to come.

additional value of products by increasing their attractiveness through enhancing the function of the products. In order to make the interpretation more easily understandable, this model supposes that IT creates a “new market", the value of which is θ. It is also assumed that since the licensee’s technology, IT, is a perfect substitute, but not a complement, firm L has a concern that firm A may “steal its market". Because of this “business stealing effect", firm L is placed at a disadvantaged position especially when the innovation is severable, which in turn lends support to the favorable effect of grant-back clauses. Lastly, since the willingness to pay in each market in the absence of IT is normalized to 1, the corresponding range of the market entry cost must be generally limited to $c \in (0, 1)$ from the above-mentioned assumption.

The licensing game is structured in what follows. In Stage 0, before the licensing game starts, firm L is in possession of BT and the attribute of the innovation, severable or non-severable, is known to both firms L and A. Although whether an innovation is severable or non-severable is not always obvious and it is sometimes judged by courts, the model assumes that the attribute of the innovation has already been determined and is common knowledge. In other words, this model eliminates all uncertainty and asymmetric information regarding the innovation to be achieved.

In Stage 1, firm L decides not only to offer firm A a contract regarding the license of BT, but also to include a grant-back clause in the contract if available. This initial contract offered by firm L is assumed to be based on a “take-it-or-leave-it offer" and that renegotiation is not allowed by the terms of the contract. Since there is a single licensor who possesses vital base technology toward follow-on innovation, firm L has total control of the use of a grant-back clause, which would reflect actual bargaining power of the firms. If an initial license contract is concluded, innovation is achieved by firm A immediately after the agreement, and subsequently, the innovation cost is realized. By contrast, if an initial license contract is not concluded, firm L serves in the both markets with BT. It is natural to assume that innovation and entry can occur only if the firms find it profitable to undertake such actions.

The important assumption is that the royalty of BT (denoted by r_1) is conditional upon usage by firm A. More precisely, when a follow-on innovation occurs, firm A is eligible to “opt out" of the royalty payment by substituting IT as long as the innovation is severable. The background logic is as follows: the innovation instantaneously occurs after technology transfer of BT, and it can effectively cancel all royalty payments accruing to BT since severable innovation does not ultimately infringe BT. Although appearing somewhat extreme, this assumption enables us to establish a clear argument of the “but for" defense of a grant-back clause.

In Stage 2, firm A decides to offer a contract regarding the license of IT in return for a royalty payment (denoted by r_2) if a grant-back clause is not exercised. When a grant-back clause is included in the contract in Stage 1, firm A cannot require firm L to pay any royalties of IT (i.e. $r_2 = 0$). [5] Subsequently, upon having completed the above two stages, firms L and A sell their products in the market and the profits are realized. See Figure 2.1 that depicts this multistage game.

[5] Put simply, this setting supposes that royalties are independent of output (indeed, output is normalized in the model), so that the form of the royalty payment (i.e. fixed or commensurate with usage) is irrelevant.

[Stage 0]

Firm L already possesses BT.
The attribute of the innovation is known to firms L and A.

[Stage 1]

Firm L decides:
- to license BT to firm A; and
- to include a grant-back clause in the license contract.

- If a license contract is concluded, innovation is immediately achieved and the game proceeds to Stage 2.
- If a license contract is not concluded, firm L enters the market of firm A.

(Innovation and entry can occur only if it is profitable to do so.)

[Stage 2]

(If the grant-back clause is not included)
Firm A decides to license IT to firm L.

The profits of firms L and A are realized.

Figure 2.1. Timing of the model (single licensee).

Let us confirm again how the consequences of the model are varied in accordance with severable and non-severable innovation, following the legal practice of the EU (2004). When the innovation is severable, firm A is eligible to sell a product with high quality, $1 + \theta$, not actually relying on BT, which is transferred from firm L. That is, severable innovation does not infringe BT and it is separately patentable. On the other hand, when the innovation is non-severable, IT cannot be used without infringing BT. Consequently, the license contract is still fully effective and firm A owes a royalty payment to firm L under non-severable innovation. Moreover, when bargaining over a license contract in Stage 1, firm L takes into account whether a grant-back clause should be included if available. When a grant-back clause is included, IT reverts back to firm L without any royalty payments, while firm A still retains the right to exploiting

IT without additional payments to firm L other than r_1. [6] When a grant-back clause is not available, firm A can still decide to grant the license of IT back to firm L while demanding a royalty rate of r_2.

It should be also noted that in almost all of the analyses presented in this chapter, territorial restrictions are applicable to any license contract that is in force. This implies that firms licensing the technologies continue to operate under territorial restrictions even if licensed firms do not actually use them. This assumption is accounted for on two grounds. First, most importantly, the territorial restrictions can be considered a simple shorthand utilized in explaining the circumstances where head-on competition between licensors and licensees are limited. For instance, Bleeke and Rahl (1979) point out the fact that firms tend to specify some form of territorial restrictions in license contracts in order to impose some restrictive circumstances. [7] Second, territorial restrictions in license contracts are, in a practical sense, usually allowed in the U.S., the EU, and Japan, as long as they do not aim at decreasing competition or contradicting competition policies through setting up a cartel. [8] Hence, we presently take it for granted that license contracts prescribe territorial restrictions which divide up markets between a licensor and a licensee. In our framework, the absence of territorial restrictions is necessarily unfavorable for a licensor because license contracts without them

[6] Exclusive grant-backs (see Footnote 3) are not assumed in this analysis. It is generally recognized that a license contract which totally prohibits licensees from using improved technologies is regarded as an unfair trade practice. See also Footnote 35 in Chapter 1.

[7] Bleeke and Rahl (1979) draw a conclusion that "$\cdots$ the response to this question strongly indicates that most corporations are not willing to compensate for the absence of restrictive [territorial] provisions by charging a higher royalty rate". We can develop an understanding that where patents are present, territorial restrictions may be sought as a prerequisite to license contracts.

[8] See the summary by Delrahim (2004) who describes positions taken by the U.S. and the EU on territorial restrictions.

inevitably create direct competitors. Afterward in Subsection 2.4.1, we investigate as an extension the case where territorial restrictions are prohibited.

Finally, the royalty rates, r_1 and r_2, are determined in each stage based on the Nash bargaining solution (hereafter, NBS). The following multi-stage game is solved by backward induction.

2.3. Baseline analysis: one potential licensee

In Section 2.3, there is a single potential licensee, firm A. We examine the incentives to license for both severable and non-severable innovation. In order to simplify the analysis, it is assumed that the innovation cost, k, is zero at the outset ($k = 0$). Even if a strictly positive innovation cost is assumed, the main result survives with only some modifications. (Subsection 2.4.2 analyzes the model of a strictly positive innovation cost.) In what follows, it is demonstrated how the “but for" defense of a grant-back clause is effective for licensing activities under severable innovation.

2.3.1. Non-severable innovation

Non-availability of grant-back clause (NN)

At the beginning, we consider the case where a grant-back clause is not available to firm L. Assuming that an initial license contract in Stage 1 has been already concluded, we determine a royalty rate in Stage 2, r_2, paid by firm L to firm A. By purchasing the license of IT, firm L can increase the value of its own market from 1 to $1 + \theta$. As a result, we obtain the NBS in Stage 2 by maximizing the following "Nash product" with regard to r_2: [9]

$$\max_{r_2} \underbrace{[(1+\theta+r_1-r_2)-(1+r_1)]}_{\text{Profit difference of Firm L}} \underbrace{[(1+\theta-r_1+r_2)-(1+\theta-r_1)]}_{\text{Profit difference of Firm A}}$$

$$= \max_{r_2}(\theta - r_2)r_2. \Rightarrow r_2^{NN} = \frac{\theta}{2}. \tag{2.1}$$

The order of the terms concerning firms L and A is from left to right. (This order is maintained hereafter.) Equation (2.1) provides the NBS, namely, $r_2^{NN} = \frac{\theta}{2}$.

Let us revert to Stage 1. By definition, since non-severable innovation always infringes BT, the royalty rate, r_1, should be paid by firm A to firm L. We represent the reservation profits of firms L and A as $\pi_L^R = 2 - c$ and $\pi_A^R = 0$, respectively, when the initial license contract is not concluded in Stage 1. Then, the initial royalty rate, r_1, is the solution of the following problem:

$$\max_{r_1}[(1+\theta+r_1-r_2^{NN})-\pi_L^R][(1+\theta-r_1+r_2^{NN})-\pi_A^R]$$

$$= \max_{r_1}\left(-1+\frac{\theta}{2}+c+r_1\right)\left(1+\frac{3\theta}{2}-r_1\right). \Rightarrow r_1^{NN} = 1+\frac{\theta}{2}-\frac{c}{2}. \tag{2.2}$$

Since $r_1^{NN} = 1+\frac{\theta}{2}-\frac{c}{2}$ from Equation (2.2), firm L earns a net royalty of $r_1^{NN}-r_2^{NN} = 1-\frac{c}{2}$. The profits firms L and A earn are formulated as $\pi_L^{NN} = 2+\theta-\frac{c}{2} > \pi_L^R$ and $\pi_A^{NN} = \theta+\frac{c}{2} > \pi_A^R$, respectively. This implies that the so-called "participation conditions"

[9] The following results are partly subject to the formulation of the symmetric Nash bargaining solution, which is consistently assumed throughout the analyses in this chapter. Alternatively, we could posit the asymmetric Nash bargaining solution of Equation (2.1), such that $\max_{r_2}[(1+\theta+r_1-r_2)-(1+r_1)]^p[(1+\theta-r_1+r_2)-(1+\theta-r_1)]^{1-p}$, where $p \in (0,1)$ denotes a relative weight for players' bargaining power. However, although the profit distribution between firms L and A depends on p, it does not seem that general tendency of the results discussed later would change by use of the asymmetric Nash bargaining solution. Hence, in order to avoid complexity, the following analyses employ the symmetric Nash bargaining solution (i.e. $p = \frac{1}{2}$).

of both firms L and A in an initial license contract are satisfied. (They are denoted as “PCL" and “PCA," respectively, for a descriptive purpose.) Accordingly, the two firms prefer to conclude a contract even when a grant-back clause is not available. We can also find that firm L benefits from the follow-on innovation achieved by firm A as its profit is increasing in θ. Although firm L needs to incur one half of the market entry cost ($\frac{c}{2}$), which is not actually realized, the entire gain of θ from IT is still retained by firm L.

Assuming that social welfare is simply viewed as a total profit of the two firms, [10] it amounts to $W^{NN} = \sum_{f \in F} \pi_f^{NN} = 2 + 2\theta$. Note that this social welfare achieves the first-best level, and therefore, the “social optimality condition" (denoted by “SC") is satisfied. Unambiguously, $W^{NN} = 2 + 2\theta$ exceeds $W^R = 2 - c$, which is obtained without any license contracts.

Availability of grant-back clause (NA)

How does the result change when firm L is allowed to include a grant-back clause into an initial license contract? As stated earlier, firm A can no longer demand a royalty in return for IT in Stage 2 due to the grant-back clause (i.e. $r_2 = 0$). The NBS of r_1 in Stage 1 is determined by the following equation:

$$\max_{r_1}[(1 + \theta + r_1) - (2 - c)][(1 + \theta - r_1) - 0]$$
$$= \max_{r_1}(-1 + \theta + c + r_1)(1 + \theta - r_1). \Rightarrow r_1^{NA} = 1 - \frac{c}{2}. \tag{2.3}$$

[10] Unlike Chapter 1, we postulate that consumers’ willingness to pay is entirely included in the profits of the two firms through extraction by firms, for example, price discrimination. This implies that the demand curve of consumers is perfectly inelastic. In this simplified case, private values of firms can be considered equivalent to social values.

Equation (2.3) provides $r_1^{NA} = 1 - \frac{c}{2}$, which is equivalent to the “net royalty" earned by firm L without any grant-backs. Not surprisingly, the profits of firms L and A, as well as social welfare, are exactly the same as in the case where a grant-back clause is not available: $\pi_L^{NA} = 2 + \theta - \frac{c}{2}$, $\pi_A^{NA} = \theta + \frac{c}{2}$, and $W^{NA} = \sum_{f \in F} \pi_f^{NA} = 2 + 2\theta$. Table 2.1 presents a summary of each case. With regard to non-severable innovation and a single licensee, we can derive the following lemma.

	Non-license	Non-availability of GB (NN)	Availability of GB (NA)
π_L	$2 - c$	$2 + \theta - \frac{c}{2}$	$2 + \theta - \frac{c}{2}$
π_A	0	$\theta + \frac{c}{2}$	$\theta + \frac{c}{2}$
W	$2 - c$	$2 + 2\theta$	$2 + 2\theta$

Note: “GB" is an abbreviation for “grant-back clause."

Table 2.1. Profits and social welfare (non-severable innovation).

Lemma 2.1 (Non-severable innovation and a single licensee) We can propose the following statements regardless of an availability of a grant-back clause for non-severable innovation:

(1) Innovation is achieved and the improved technology is shared by firms L and A;

(2) First-best social welfare is generated; and

(3) A grant-back clause has no effects on profits or on social welfare.

Since an initial license contract is essential to facilitating IT, both firms L and A can potentially benefit from it. The reason why benefits of firm L are produced by licensing is that even without a grant-back clause, the initial royalty rate, r_1, is determined so as to capture future gains from an innovation resulting from the anticipated follow-on licensing activity. More generally, the net royalty, $r_1^{NN} - r_2^{NN}$, earned by firm

L can be established to the same level as r_1^{NA} by adjusting r_1^{NN} in an appropriate manner. This is why these two ways of capturing the outcome of the innovation should be equivalent; however, the only observable difference is a one-way or two-way flow of royalty payments between the two firms.

2.3.2. Severable innovation

Under severable innovation, firm L cannot extract any royalty payment from firm A in return for transferring BT in Stage 1. Repeatedly, this argument is built on a simple assumption: follow-on innovation immediately makes BT obsolete, and hence, although firm A uses BT as a tool to innovate, firm A neither actually employs this technology in production, nor infringes it. It follows that after granting a license in Stage 1, firm L holds BT only in accordance with the value of 1 while firm A holds IT in accordance with the value of $1 + \theta$. Furthermore, it is reasonable to suppose that in light of the attribute of severable innovation, the territorial restriction still applies to the “original market" (the value of which is 1), but not to the “new market" (the value of which is θ) created by the innovation. Consequently, firm A can potentially earn $\theta - c$ (if θ is greater than c) in the new market of firm L in addition to $1 + \theta$ in its own market.

Non-availability of grant-back clause (SN)

Whether firm A enters the new market of firm L depends on the value of the innovation (θ) relative to the market entry cost (c). In the first place, suppose that the innovation is relatively large (i.e. $\theta > c$), so that firm A is likely to enter the new market of firm L. Given that firm L issues a license of BT in Stage 1, we obtain the NBS in Stage 2 in the following equation:

$$\max_{r_2}[(1+\theta-r_2)-1][(\underbrace{1+\theta}_{\text{Market A}}+r_2)-(\underbrace{1+\theta}_{\text{Market A}}+\underbrace{\theta-c}_{\text{Market L}})]$$

$$=\max_{r_2}(\theta-r_2)(-\theta+c+r_2).\Rightarrow r_2^{SN}=\theta-\frac{c}{2}. \tag{2.4}$$

It is notable that firm L cannot demand a royalty payment in Stage 1 for severable innovation. Hence, the profits of firms L and A are $\pi_L^{SN} = 1+\frac{c}{2} > 1$ and $\pi_A^{SN} = 1+2\theta-\frac{c}{2} > 1+2\theta-c$, respectively. Consequently, firms L and A agree with the application of the license of IT. But we also need to check whether firm L intends to grant the license of BT in Stage 1. Comparing $\pi_L^{SN} = 1+\frac{c}{2}$ with $\pi_L^R = 2-c$, we can easily see that PCL is satisfied if $\pi_L^{SN} > \pi_L^R \Leftrightarrow \frac{2}{3} < c < 1$. In other words, if the market entry cost is sufficiently large, firm L grants BT to firm A. The reason is that when the market entry cost is large, firm A can retain a relatively low profit by entry in Stage 2 ($\pi_A^{SN} = 1+2\theta-\frac{c}{2}$), and on the other hand, the profit of firm L increases ($\pi_L^{SN} = 1+\frac{c}{2}$) in contrast to that in the absence of the license of BT. As a result, firm L does have a clear motive to grant BT.

Next, when the innovation is relatively small (i.e. $\theta < c$), firm A will not choose to enter a new market of firm L. The NBS in Stage 2 is provided by:

$$\max_{r_2}[(1+\theta-r_2)-1][(1+\theta+r_2)-(1+\theta)]=\max_{r_2}(\theta-r_2)r_2.\Rightarrow r_2^{SN}=\frac{\theta}{2}. \tag{2.5}$$

The profits of firms L and A are calculated as $\pi_L^{SN} = 1+\frac{\theta}{2}$ and $\pi_A^{SN} = 1+\frac{3\theta}{2}$, respectively. Comparing the profits of firm L between $\pi_L^{SN} = 1+\frac{\theta}{2}$ and $\pi_L^R = 2-c$, we can see that PCL is satisfied if $\pi_L^{SN} > \pi_L^R \Leftrightarrow \theta > 2(1-c)$. This indicates that $\frac{2}{3} < c < 1$ is absolutely necessary. [11] (By contrast, if $0 < c < \frac{2}{3}$ holds, PCL is never satisfied.) The intuition for the condition, $\theta > 2(1 - c)$, is straightforward. That is, although the profit of firm L ($\pi_L^{SN} = 1 + \frac{\theta}{2}$) does not directly depend on a market entry cost, c, its profit earned by an entry into the market of firm A without an initial license contract ($\pi_L^R = 2 - c$) will be decreased as the market entry cost gets large. This makes firm L recognize that if the value of innovation, θ, is sufficiently large relative to the market entry cost, granting BT to firm A and encouraging innovation is more profitable.

Figure 2.2 indicates the shaded area of (c, θ) where both PCL and PCA are satisfied (which means that an initial license contract is possible), [12] and thereby, social welfare attains the first-best level (SC is satisfied), $W^{SN} = \sum_{f \in F} \pi_f^{SN} = 2 + 2\theta$.

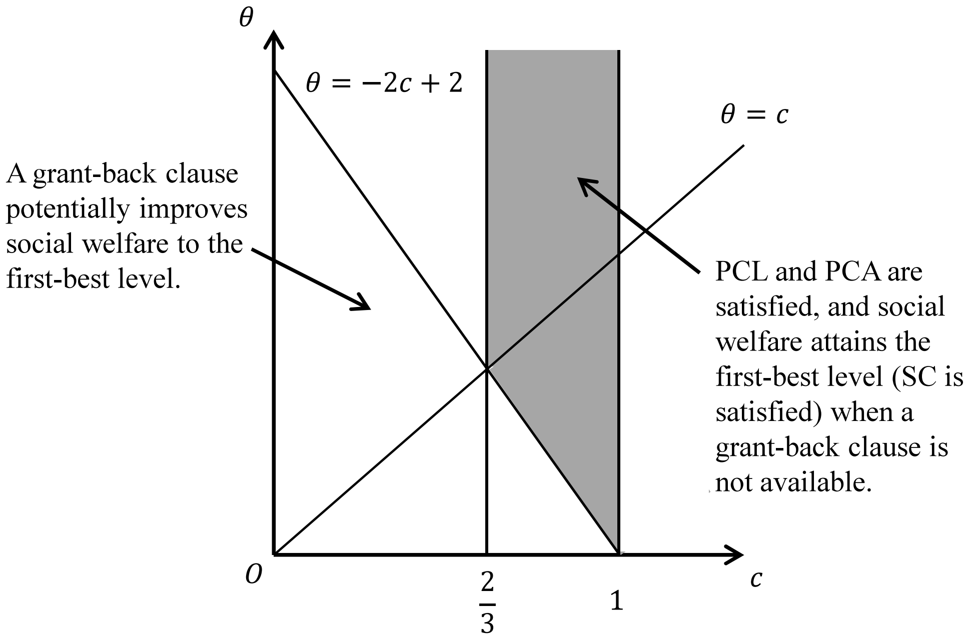


Figure 2.2. Diagram of (c, θ) (severable innovation).

[11] Because $\theta < c$, the relation $2 - 2c < \theta < c$ provides $\frac{2}{3} < c < 1$.
[12] PCA is obviously satisfied for every θ and c.

Lemma 2.2 (Severable innovation and a single licensee) Suppose that a grant-back clause is not available. We can propose the following statement:

(1) Innovation is not achieved for $0 < c < \frac{2}{3}$ regardless of the value, θ; and

(2) Innovation is achieved, and therefore, first-best social welfare is generated if both $\frac{2}{3} < c < 1$ and $\theta > 2(1 - c)$ hold.

Availability of grant-back clause (SA)

We need to note that when a grant-back clause is used, firm L earns $\pi_L^{SA} = 2 + \theta - \frac{c}{2}$ regardless of whether the innovation is severable or non-severable. (See the previous discussion in NA.) Despite the severable innovation, the grant-back clause still enables firm L (a licensor) to make firm A (a licensee) pay a royalty for BT and to be exempted from making her own royalty payment for IT through the patent right accruing to such a clause. The profit of firm L is compared between π_L^{SA} and π_L^{SN} as summarized in Table 2.2. It can be demonstrated that if firm L has total control of whether to include a grant-back clause in the license contract, the grant-back contract will be the preferred conduct for firm L irrespective of the relation between θ and c. [13]

[13] We need to take into account the assumption that firm L provides a take-it-or-leave-it offer and renegotiation is eliminated. Since the total bargaining power over including a grant-back clause lies on the side of firm L, firm A is compelled to accept the offer as long as PCA is satisfied. If we accept the situation where firm A has the right to refuse a grant-back clause, the following argument would no longer be feasible.

	Non-license	Non-availability of GB (SN)		Availability of GB (SA)
		$\theta > c$	$\theta < c$	
π_L	$2 - c$	$1 + \frac{c}{2}$	$1 + \frac{\theta}{2}$	$2 + \theta - \frac{c}{2}$
π_A	0	$1 + 2\theta - \frac{c}{2}$	$1 + \frac{3\theta}{2}$	$\theta + \frac{c}{2}$
W	$2 - c$	$2 + 2\theta$	$2 + 2\theta$	$2 + 2\theta$

Table 2.2. Profits and social welfare (severable innovation).

Proposition 2.1 (Severable innovation and a single licensee)

(1) Firm L prefers to include a grant-back clause in an initial license contract;

(2) Firm A accepts it and innovation is necessarily achieved; and

(3) A grant-back clause improves social welfare to the first-best level when $\frac{2}{3} < c < 1$ and $\theta > 2(1 - c)$ are not satisfied.

Proof (1) With regard to $\theta > c$, $\pi_L^{SA} - \pi_L^{SN} = (2 + \theta - \frac{c}{2}) - (1 + \frac{c}{2}) = 1 + (\theta - c) > 0 \Leftrightarrow \pi_L^{SA} > \pi_L^{SN}$. In addition, with regard to $\theta < c$, $\pi_L^{SA} - \pi_L^{SN} = (2 + \theta - \frac{c}{2}) - (1 + \frac{\theta}{2}) = (1 - \frac{c}{2}) + \frac{\theta}{2} > 0 \Leftrightarrow \pi_L^{SA} > \pi_L^{SN}$ (because $c \in (0, 1)$ is assumed). Hence, $\pi_L^{SA} > \pi_L^{SN}$ holds for every (c, θ).

(2) It is quite obvious that $\pi_A^{SA} = \theta + \frac{c}{2} > \pi_A^R = 0$.

(3) When $\frac{2}{3} < c < 1$ and $\theta > 2(1 - c)$ are not satisfied, innovation is not achieved and social welfare remains as $W^R = 2 - c$. However, a grant-back clause brings about innovation for every (θ, c), and thereby, $W^{SA} = 2 + 2\theta$ is larger than $W^R = 2 - c$. ■

The implication of Proposition 2.1 is clear; innovation is always achieved and social welfare attains the first-best level through a grant-back clause even for severable innovation, where firm L is originally placed in a more disadvantaged position than non-severable innovation regarding the royalty of BT. This proposition overturns Lemma 2.2 which indicated that an initial license contract may not be concluded in some configurations of (c, θ) with no grant-backs, as illustrated by a non-shaded area in Figure 2.2. It is now represented as a potential configuration where a grant-back clause can improve social welfare to the first-best level. On the other hand, although the profit of firm A is still greater than zero (which is the level obtained in the absence of the license of BT), it is subject to a decline due to the grant-back clause. [14] This suggests that since first-best social welfare is also attained by a grant-back clause, the surplus is redistributed from firm A to firm L along with a grant-back. Hereby, even when PCL is not satisfied for severable innovation, a contract with a grant-back clause enables firm L to grant BT to firm A, and at the same time, to encourage firm A to further innovate. Accordingly, social welfare is recovered to the first-best level.

It is worthwhile making a comparison with the insight provided by Scotchmer (1991, 1996). She investigates the incentive of a second inventor to further innovate in cases where bargaining is conducted *ex post* and *ex ante* innovation, respectively. In her model, the problem with *ex post* bargaining is that since sunk innovation costs cannot be taken into account in a license negotiation, a licensee is subject to “hold up". If a licensee is not sufficiently compensated for innovation costs, socially beneficial innovation led by the licensee may not occur in her framework. In this regard, Scotchmer (1991, 1996) proposes that *ex ante* bargaining can successfully solve the difficulty in maintaining the licensee’s incentive, as a licensor makes a commitment to

[14] With regard to $\theta > c$, $\pi_A^{SA} - \pi_A^{SN} = (\theta + \frac{c}{2}) - (1 + 2\theta - \frac{c}{2}) = (c - 1) - \theta < 0 \Leftrightarrow \pi_A^{SA} < \pi_A^{SN}$. In addition, with regard to $\theta < c$, $\pi_A^{SA} - \pi_A^{SN} = (\theta + \frac{c}{2}) - (1 + \frac{3\theta}{2}) = (\frac{c}{2} - 1) - \frac{\theta}{2} < 0 \Leftrightarrow \pi_A^{SA} < \pi_A^{SN}$. Hence, $\pi_A^{SA} < \pi_A^{SN}$ holds for every (θ, c).

paying up front the innovation cost.

As discussed before, this chapter finds that if the innovation is severable, it is expected to occur with a grant-back clause, but may not without it. The underlying mechanism is that whereas a royalty rate is set *ex ante* innovation by a licensor when a grant-back clause is available, it is eventually set *ex post* innovation by a licensee when a grant-back clause is not available. The mechanism presented here is entirely different from that of Scotchmer (1991, 1996). More precisely, while the *ex post* bargaining deteriorates the incentive of a licensee through a “hold-up" problem in Scotchmer’s model, it does that of a licensor under severable innovation in our model. Additionally, although her model requires a positive innovation cost for the analysis to be applicable, this chapter refers to the possibility that even if an innovation cost is set at zero, socially efficient follow-on innovation may not evolve without grant-back clauses for severable innovation. The reason why the incentive of a licensor falls short of the socially optimal level for severable innovation is that a licensor is threatened in the first stage by a licensee, who may potentially avoid an *ex ante* royalty payment for the base technology and may erode the licensor’s own new market. Such prior concerns held by a licensor culminate in her rescinding the initial license contract, which eventually leads to no follow-on innovation achieved by a licensee.

To sum up, a grant-back clause not only guarantees that a licensor is entitled to receiving a royalty payment from a licensee in the first stage, but also prevents a licensee’s entry into the licensor’s new market, which can provide a licensor with an incentive to transfer her base technology in severable innovation. While this analysis is closely related to the “but for" defense of a grant-back clause in terms of a licensor’s innovation incentive as regards to severable innovation, the flavor of superiority of *ex ante* bargaining still exists as demonstrated by Scotchmer (1991, 1996).

2.4. Extension

In order to confirm the robustness of our analysis that suggests the effectiveness of a grant-back clause, the model is extended in the two directions. The first case prohibits territorial restrictions by a competition authority, and the second case includes a strictly positive innovation cost. We can determine if the “but for" defense of grant-back clauses is still an essential element to spur follow-on innovation which is severable in these two cases.

2.4.1. Prohibition of territorial restrictions

This subsection is a brief attempt to investigate how eliminating the assumption of territorial restrictions affects the results discussed so far. This analysis is noteworthy for reflecting the recent tendency in competition policies that have attached more importance to the total prohibition of territorial restrictions. [15]

Four cases are reviewed as a whole: “Non-severable innovation and non-availability

[15] Until recently, the attitude of the EU toward territorial restrictions in license contracts was fairly relaxed. In a nutshell, European antitrust law generally tolerated agreements that prevent a licensee from competing in the market of a licensor or equally a licensor from competing in the market of a licensee. However, the European Commission has recently become more concerned about territorial restrictions because they may decrease positive economic effects induced by competition. For example, see Wäktare (2007), who surveys a territorial restrictions case in gas contracts regarding the import of Algerian gas into Europe.

of a grant-back clause" (Case 1: NN); “Non-severable innovation and availability of a grant-back clause" (Case 2: NA); “Severable innovation and non-availability of a grant-back clause" (Case 3: SN); and “Severable innovation and availability of a grant-back clause" (Case 4: SA). See Table 2.3 for these four cases.

	Non-availability of GB	Availability of GB
Non-severable innovation	Case 1: NN	Case 2: NA
Severable innovation	Case 3: SN	Case 4: SA

Table 2.3. Attributes of innovation and the use of a grant-back clause.

■ **Case 1: NN**

In order to facilitate the later analysis, we place some assumptions on a territorial restriction. First, if the territorial restriction between firms L and A is invalid, firm A can unilaterally enter both the original and new markets of firm L after achieving innovation with the use of IT. Second, it is reasonable to think that firm A, which develops IT, possesses relatively a competitive advantage over firm L in both of these markets, because firm A is considered to be more familiar with BT and IT than firm L through learning in the innovation process. [16] Accordingly, when IT is licensed to firm L, the total market scale of firm L is assumed to decrease from $1 + \theta$ to $x(1 + \theta)$ with $x \in (0, 1)$, due to the direct market competition against firm A. The parameter x denotes the degree of disadvantage firm L faces in a competitive position relative to firm A. On the other hand, firm A is assumed to have obtained the market of firm L, of which the value is $(1 - x)(1 + \theta)$. It is also posited that if firm L is not the recipient

[16] Since firm A needs to learn much about BT in innovating IT, it is assumed to be more familiar with BT as well as IT than firm L.

of IT, it earns only x from the original market of firm L, but firm A can earn a total of $1 - x + \theta$ from both original $(1 - x)$ and new (θ) markets of firms L. In other words, whereas direct competition occurs in the original market of firm L, firm A is placed in an exclusively advantageous position in the new market. Moreover, for the purpose of analytical simplicity, this subsection temporarily assumes that revenue obtained from entry is always larger than the cost (i.e. $1 > 1 - x > c$ and $\theta > x\theta > c$).

Based on these settings, the NBS in Stage 2 is provided by:

$$
\begin{aligned}
&\max_{r_2} \underbrace{[x(1+\theta) + r_1 - r_2 - (x + r_1)]}_{\text{Profit difference of firm L}} \\
&\underbrace{[\underbrace{1+\theta}_{\text{Market A}} + \underbrace{(1-x)(1+\theta)}_{\text{Market L}} - 2c - r_1 + r_2\} - \{\underbrace{1+\theta}_{\text{Market A}} + \underbrace{(1-x+\theta)}_{\text{Market L}} - 2c - r_1]}_{\text{Profit difference of firm A}} \\
&= \max_{r_2} (x\theta - r_2)(-x\theta + r_2). \Rightarrow r_2^{NN} = x\theta. \qquad (2.6)
\end{aligned}
$$

Moving back to Stage 1, we determine the NBS in Stage 1 as follows:

$$
\begin{aligned}
&\max_{r_1} [x(1+\theta) + r_1 - r_2^{NN} - (2-c)][1 + \theta + (1-x)(1+\theta) - 2c - r_1 + r_2^{NN} - 0] \\
&= \max_{r_1} (-2 + x + c + r_1)(2 - x + 2\theta - 2c - r_1). \Rightarrow r_1^{NN} = 2 + \theta - \frac{3c}{2}. \qquad (2.7)
\end{aligned}
$$

From these solutions, the profits are $\pi_L^{NN} = 2 + \theta - \frac{3c}{2}$ and $\pi_A^{NN} = \theta - \frac{c}{2}$, respectively. In addition, social welfare amounts to $W^{NN} = \sum_{f \in F} \pi_f^{NN} = 2 + 2\theta - 2c$.

Surprisingly, the profit of firm A remains at only $\theta - \frac{c}{2}$, which is exactly the same as the profit obtained with the territorial restriction. This means that even when firm A is eligible to enter both the original and new markets of firm L, it cannot earn a higher profit than what is projected. The reason for this is that firm L imposes a high royalty on firm A in Stage 1, anticipating that firm A will invade its markets and will deprive the potential consumers that firm L is expected to retain. Although firm L still keeps the incentive to provide an initial license of BT, [17] its profit is reduced because of the market entry by firm A. [18] Additionally, since there is always a market entry and associated competition between the firms, social welfare, $W^{NN} = 2 + 2\theta - 2c$, is smaller than the first-best level, $2 + 2\theta$, as the market entry cost, $2c$, is deducted.

The result critically relies on the setting. This model abstracts from a "positive competition effect" such as an increase in the innovation value (the willingness to pay of consumers, θ) induced, for example, by a decrease in the price of products. Nevertheless, we can say that prohibiting territorial restrictions is undesirable for non-severable innovation in the absence of a grant-back clause unless competition between firms manages to produce some positive economic effects. (See the discussion of positive competition effects in the supplementary note.)

■ **Case 2: NA**

Even if territorial restrictions are prohibited, a grant-back clause does not allow firm A to exploit IT in the markets of firm L because the patent right exclusively accrues to firm L by the nature of that clause. However, firm A is eligible to enter the original market of firm L, being more accustomed with BT than firm L as assumed before. [19]

---

[17] $\pi_L^{NN} = 2 + \theta - \frac{3c}{2} > \pi_L^R = 2 - c$ because $\pi_L^{NN} - \pi_L^R = \theta - \frac{c}{2} > 0$ for $\theta > c$.

[18] Indeed, these profits do not depend on what assumptions we make about a market entry. Even if it is assumed that firm L can earn 0, instead of $x \in (0, 1)$, within its own market, the profits of firms L and A also become $\pi_L^{NN} = 2 + \theta - \frac{3c}{2}$ and $\pi_A^{NN} = \theta - \frac{c}{2}$, respectively.

[19] Despite the prohibition of territorial restrictions, firm L is assumed not to enter the new market of firm A to simplify this analysis. We can think that since firm L is less familiar with IT than firm A, it might lose an incentive to compete in this market.

The problem in Stage 1 is modified as:

$$\max_{r_1}[(1 - x + \theta + r_1) - (2 - c)][(1 + \theta + x - r_1) - 0]$$

$$= \max_{r_1}(-1 - x + \theta + c + r_1)(1 + x + \theta - r_1). \Rightarrow r_1^{NA} = 1 + x - c. \tag{2.8}$$

The profits of firms L and A are $\pi_L^{NA} = 2 + \theta - c$ and $\pi_A^{NA} = \theta$, respectively. Since we derive $\pi_L^{NA} = 2 + \theta - c > \pi_L^{NN} = 2 + \theta - \frac{3c}{2}$, firm L is certain to include a grant-back clause into the contract if available. That is how grant-back clauses play a significant role in restoring social welfare to $W^{NA} = \sum_{f \in F} \pi_f^{NA} = 2 + 2\theta - c$, by potentially saving a market entry cost for firm A seeking entry into the new market of firm L. And yet, since firm A actually enters the original market of firm L at a cost of c, social welfare is deducted by this cost from the first-best level, $2 + 2\theta$. Even so, we can see that the inclusion of a grant-back clause is desirable even for non-severable innovation.

■ **Case 3: SN**

The NBS in Stage 2 is provided by:

$$\max_{r_2}[\{x(1 + \theta) - r_2\} - x]$$

$$[\{(1 + \theta) + (1 - x)(1 + \theta) - 2c + r_2\} - \{(1 + \theta) + (1 - x + \theta) - 2c\}]$$

$$= \max_{r_2}(x\theta - r_2)(-x\theta + r_2). \Rightarrow r_2^{SN} = x\theta. \tag{2.9}$$

The profits of firms L and A in terms of Stage 2 are $\tilde{\pi}_L^{SN} = x$ and $\tilde{\pi}_A^{SN} = 2 - x + 2\theta - 2c$, respectively. In view of severable innovation, prohibiting territorial restrictions results in a severe consequence for firm L, which can earn a small positive profit ($x < 1$). Unsurprisingly, firm L will never conclude an initial license contract with firm A, because the license of BT creates a strong competitor without any expectation that firm L can charge and receive a royalty payment from firm A. Unlike the previous analyses of Cases 1 and 2, there is not even the slightest possibility that innovation will be achieved within the parameters of (c, θ) because $\tilde{\pi}_L^{SN} = x < \pi_L^R = 2 - c$. [20] Consequently, the profits of firms L and A boil down to $\pi_L^{SN} = \pi_L^R = 2 - c$ and $\pi_N^{SN} = \pi_A^R = 0$, respectively, in the absence of a license contract. In addition, social welfare achieves the lowest level, that is, $W^{SN} = \sum_{f \in F} \pi_f^{SN} = 2 - c$.

■ **Case 4: SA**

Similar to the analysis in Section 2.3, the results are identical to Case 2 due to the effectiveness of a grant-back clause even for severable innovation: by enforcing a royalty payment for BT and nullifying it for IT. At the end, grant-back clauses enable innovation to be realized and social welfare to be restored to the second-best level, $W^{SA} = \sum_{f \in F} \pi_f^{SA} = 2 + 2\theta - c$.

Proposition 2.2 (Prohibition of territorial restrictions and a single licensee)

(1) Even if a grant-back clause is not available for non-severable innovation, innovation is achieved and the improved technology is shared by firms L and A. However, first-best social welfare is not generated;

(2) If a grant-back clause is not available for severable innovation, innovation is not achieved so that the lowest social welfare is generated; and

(3) A grant-back clause allows social welfare to restore the second-best level for both non-severable and severable innovation.

---

[20] $\pi_L^R - \tilde{\pi}_L^{SN} = 2 - x - c > 1 - c > 0 \Leftrightarrow \pi_L^R > \tilde{\pi}_L^{SN}$ because $0 < c < 1$ and $0 < x < 1$ are assumed.

The profits and social welfare are presented in Table 2.4. Indeed, severable innovation without territorial restrictions causes firm L to entirely lose the incentive to provide the license of BT. But the same mechanism works as before in the use of a grant-back clause, so that it recovers the incentive of firm L to transfer BT for severable innovation. Moreover, although there is neither an advantage nor disadvantage of possessing a grant-back clause in non-severable innovation in the case where territorial restrictions are permitted, it contributes to the improvement in social welfare when territorial restrictions are prohibited. More generally, the grant-back avoids the welfare loss that can be potentially caused by the threat of direct competition in a new market under non-severable innovation. Hence, a stronger stance can be taken in support of the “but for" defense of a grant-back clause to encompass both severable and non-severable innovation.

	Case 1: NN	Case 2: NA	Case 3: SN	Case 4: SA
π_L	$2+\theta-\frac{3c}{2}$	$2+\theta-c$	$2-c$	$2+\theta-c$
π_A	$\theta-\frac{c}{2}$	θ	0	θ
W	$2+2\theta-2c$	$2+2\theta-c$	$2-c$	$2+2\theta-c$

Table 2.4. Profits and social welfare (prohibition of territorial restrictions).

Supplementary note

The “but for" defense of a grant-back clause with regard to non-severable innovation discussed above is based on the assumption that direct competition between firms L and A never creates additional value to innovation that follows. Instead, suppose that the innovation value, θ, increases to $\hat{\theta} > \theta$ through competition under non-severable innovation when a grant-back clause is not used (that is, a positive competition effect is generated). The possible ground for this is that the firms intend to win over their competitors in a new market by creating attractive products using enhanced IT. In this situation, social welfare also increases to $\hat{W}^{NN} = 2 + 2\hat{\theta} - 2c$.

Comparing social welfare between $\hat{W}^{NN}$ and W^{NA}, we find that the use of a grant-back clause is desirable provided that $W^{NA} = 2 + 2\theta - c > \hat{W}^{NN} = 2 + 2\hat{\theta} - 2c \Leftrightarrow c > 2(\hat{\theta} - \theta)$. This result implies that saving of a market entry cost through a grant-back (c) overwhelms the gain of a positive competition effect from the new markets of both firms L and A ($2(\hat{\theta} - \theta)$). It should be also noted that firm L will utilize a grant-back clause under the same condition of the above, namely $\hat{\pi}_L^{NA} = 2 + \hat{\theta} - c > \pi_L^{NN} = 2 + \theta - \frac{3c}{2} \Leftrightarrow c > 2(\hat{\theta} - \theta)$. Hence, these findings can be summarized as follows.

Lemma 2.3 Suppose that competition between firms L and A due to the prohibition of territorial restrictions enhances the innovation value from θ to $\hat{\theta}$ in the new markets of firms L and A. When innovation is non-severable, we can obtain:

(1) A grant-back clause is socially desirable in generating second-best social welfare if a market entry cost is larger than an increase in the total innovation value (i.e. $c > 2(\hat{\theta} - \theta)$); and

(2) Firm L concludes a license contract that includes a grant-back clause if and only if it improves social welfare.

2.4.2. Positive innovation cost

This subsection examines how the “but for" defense should be modified when an innovation cost incurred by a licensee is strictly positive (i.e. $k > 0$). The timing of the contract is crucial in this analysis: whether an innovation cost is realized before or after the bargaining process. In order to obtain explicit results, this chapter assumes that an innovation cost arises *ex ante* Stage 2 (license of IT), and at the same time as Stage 1 (development of IT), which seems to be in line with the common sense. This suggests that firm A is unable to recover the development cost of IT when concluding the license of IT, that is, a hold-up problem occurs. Under this setting, we not only derive a result that is analogous to Proposition 2.1, but also demonstrate that the “but for" defense of a grant-back clause for severable innovation can be applied to the recovery of the incentive of a licensee. As mentioned above, the result that bargaining *ex ante* follow-on innovation benefits a licensee by mitigating a hold-up problem is an essential point that is investigated by Scotchmer (1991, 1996).

■ Case 1: NN

Since firm A incurs an innovation cost that has been realized before Stage 2, its reservation profit also covers it. Accordingly, the NBS in Stage 2 is provided by:

$$\begin{aligned}&\max_{r_2}[(1+\theta+r_1-r_2)-(1+r_1)][(1+\theta-k-r_1+r_2)-(1+\theta-k-r_1)]\\&=\max_{r_2}(\theta-r_2)r_2.\Rightarrow r_2^{NN}=\frac{\theta}{2}.\end{aligned}\tag{2.10}$$

It is not surprising that Equation (2.10) reduces to Equation (2.1) (derived in the problem with the zero innovation cost in Subsection 2.3.1) because k is canceled out. As a result, the NBS, $r_2^{NN} = \frac{\theta}{2}$, is not contingent on k.

Referring back to Stage 1, we obtain the NBS in Stage 1 from:

$$\max_{r_1}[(1 + \theta + r_1 - r_2^{NN}) - (2 - c)][(1 + \theta - k - r_1 + r_2^{NN}) - 0]$$
$$= \max_{r_1}\left(-1 + \frac{\theta}{2} + c + r_1\right)\left(1 + \frac{3\theta}{2} - k - r_1\right). \Rightarrow r_1^{NN} = 1 + \frac{\theta}{2} - \frac{c + k}{2}. \qquad (2.11)$$

Note that if k is large enough that $k > 2 + \theta - c$ holds, $r_1^{NN} < 0$ and the actual flow of the payment would be from firm L to firm A. The solution implies that royalty revenue of firm L is decreased by the innovation cost, too. From these, the profits are $\pi_L^{NN} = 2 + \theta - \frac{c+k}{2}$ and $\pi_A^{NN} = \theta + \frac{c-k}{2}$, respectively, which suggests that the innovation cost is shared equally by the two firms. In addition, social welfare amounts to $W^{NN} = \sum_{f\in F} \pi_f^{NN} = 2 + 2\theta - k$.

PCL and PCA (which denote the participation conditions of firms L and A, respectively) require that $\pi_L^{NN} > \pi_L^R = 2 - c$ and $\pi_A^{NN} > \pi_A^R = 0$, respectively. From the two relations, a common condition such as $k < 2\theta + c$ can be derived. Moreover, social welfare is improved if SC (social optimality condition) is satisfied, that is, $W^{NN} > W^R = 2-c$. It can be easily demonstrated that this condition is reduced exactly to $k < 2\theta + c$. The participation and social optimality conditions can be interpreted as expressing that an innovation cost is smaller than the total value of innovation including the opportunity cost (that is, the market entry cost). Consequently, since PCL, PCA, and SC are all equivalent, a series of the license contracts benefit social welfare if and only if they are all concluded. Figure 2.3 depicts these conditions in the diagram of (c, θ).

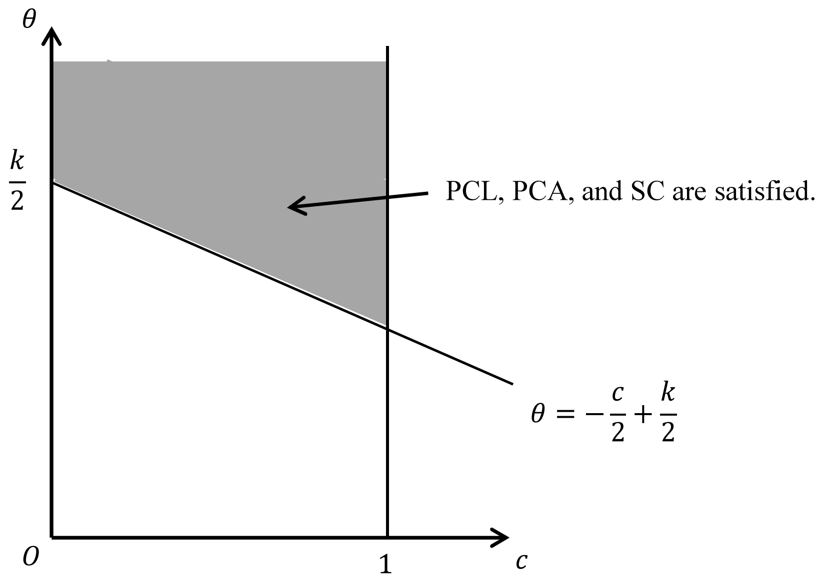


Figure 2.3. Diagram of (c, θ) (non-severable innovation; positive innovation cost).

■ **Case 2: NA**

Stage 2 is omitted because of the grant-back clause. In this case, the NBS in Stage 1 is provided by:

$$\max_{r_1}[(1 + \theta + r_1 - (2 - c)][(1 + \theta - k - r_1) - 0]$$
$$= \max(-1 + \theta + c + r_1)(1 + \theta - k - r_1). \Rightarrow r_1^{NA} = 1 - \frac{c + k}{2}. \tag{2.12}$$

The profits and social welfare are summarized as: $\pi_L^{NA} = 2 + \theta - \frac{c+k}{2}$, $\pi_A^{NA} = \theta + \frac{c-k}{2}$, and $W^{NA} = 2 + 2\theta - k$, respectively. It is clearly shown that $\pi_L^{NA} = \pi_L^{NN}$, $\pi_A^{NA} = \pi_A^{NN}$, and $W^{NA} = W^{NN}$. As a matter of course, PCL, PCA, and SC are all reduced to $k < 2\theta + c$. Under a strictly positive innovation cost, we can accordingly reconfirm the essence of Lemma 2.1 which states that social welfare can be improved through a license contract regardless of availability of a grant-back clause under non-severable innovation. But there exists a slight distinction from Lemma 2.1: since PCL, PCA, and SC are totally equivalent in the current analysis, the decision of the two firms always leads to optimal social welfare irrespective of whether innovation occurs or not. This condition is also exhibited in Figure 2.3 comprising the same diagram of (c, θ).

■ **Case 3: SN**

As analyzed previously, it is necessary to separate the cases in accordance with the relation between θ and c. First, consider the case $\theta > c$ where firm A can afford to enter the new market of firm L. The NBS in Stage 2 is provided by:

$$\max_{r_2}[(1+\theta-r_2)-1][(1+\theta-k+r_2)-(1+\theta-k+\theta-c)]$$
$$= \max_{r_2}(\theta-r_2)(-\theta+c+r_2). \Rightarrow r_2^{SN} = \theta - \frac{c}{2}. \tag{2.13}$$

Stage 1 (where firms L and A conclude a payment contract regarding the license of BT) is omitted due to the characteristics of severable innovation. This case assumes that neither can firm L require a royalty payment for BT, nor can firm A request firm L to inherit an innovation cost of IT in Stage 1. That is, there are no links between BT and IT; since IT is separated from BT, firm A cannot require the sharing of an innovation cost either. The profits and social welfare are summarized as: $\pi_L^{SN} = 1 + \frac{c}{2}$, $\pi_A^{SN} = 1 + 2\theta - \frac{c}{2} - k$, and $W^{SN} = 2 + 2\theta - k$, respectively. From these results, PCL, PCA, and SC are represented by $\frac{2}{3} < c < 1$, $k < 1 + 2\theta - \frac{c}{2}$, and $k < 2\theta + c$, respectively.

Next, suppose that firm A cannot afford to enter the market of firm L, namely, $\theta < c$. The NBS in Stage 2 is provided by:

$$\max_{r_2}[(1+\theta-r_2)-1][(1+\theta-k+r_2)-(1+\theta-k)]$$

$$= \max_{r_2}(\theta - r_2)r_2. \Rightarrow r_2^{SN} = \frac{\theta}{2}. \tag{2.14}$$

Likewise, the profits and social welfare are summarized as: $\pi_L^{SN} = 1 + \frac{\theta}{2}$, $\pi_A^{SN} = 1 + \frac{3\theta}{2} - k$, and $W^{SN} = 2 + 2\theta - k$, respectively. In turn, PCL, PCA, and SC are represented by $\theta > 2(1-c)$, $k < 1 + \frac{3\theta}{2}$, and $k < 2\theta + c$, respectively. (See Table 2.5.)

	PCL	PCA	SC
$\theta > c$	$\frac{2}{3} < c < 1$	$k < 1 + 2\theta - \frac{c}{2}$	$k < 2\theta + c$
$\theta < c$	$\theta > 2(1-c)$	$k < 1 + \frac{3\theta}{2}$	$k < 2\theta + c$

Table 2.5. Summary of PCL, PCA, and SC in Case 3.

We need to be reminded that both PCL and PCA must be satisfied in Stage 1 in order to achieve the follow-on innovation. PCL is totally the same as the basic analysis that was discussed in Section 2.3. As Lemma 2.2 discusses, firm L provides the license of BT if $\frac{2}{3} < c < 1$ (for $\theta > c$) and $\theta > 2(1-c)$ (for $\theta < c$) are satisfied. But on the other hand, firm A does not now accept the initial license contract unless $k < 1+2\theta-\frac{c}{2}$ (for $\theta > c$) and $k < 1 + \frac{3\theta}{2}$ (for $\theta < c$) are satisfied. In other words, the innovation cost must be sufficiently small that firm A engages in the innovation process. SC also needs to be satisfied in order to improve social welfare through the innovation.

In the first place, it can be clearly demonstrated that PCA is always satisfied for every (c, θ) when $0 < k \leq 1$ holds. In this case, for the purpose of checking the participation conditions, it is sufficient to focus solely on PCL. Figure 2.4 depicts a diagram in the case of $k = 1$ for descriptive simplicity. The shaded area, $ABCD$, denotes the set of (c, θ) where both PCL and PCA are satisfied. We can find that when these conditions are satisfied, SC is also satisfied. More concretely, this means that if firm L makes an initial license contract and firm A accepts it, social welfare is always improved. And yet, there remains the horizontal stripes area, $ABCEF$, that locates above the line of $\theta = -\frac{c}{2} + \frac{k}{2}$. In this area, PCL is not satisfied while SC is satisfied, so that they are inconsistent.

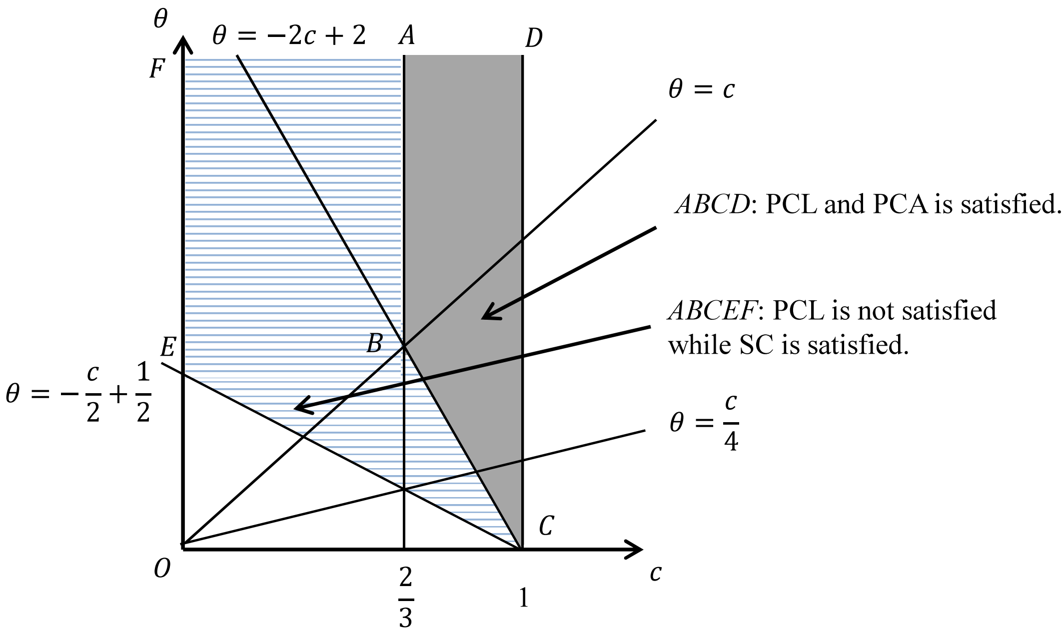


Figure 2.4. Diagram of (c, θ) (severable innovation; $k = 1$).

When $k > 1$ holds, an even more complcated situation arises. Figure 2.5 illustrates the conditions, for instance, in the case of $k = 2$. The shaded area, $ABHD$, denotes the set of (c, θ) where PCL and PCA are satisfied, and this area always satisfies SC. However, in addition to the inefficient horizontal stripes area, $ABEF$ (PCL is not satisfied while SC is satisfied), we have another type of inefficient area. Specifically, the meshed area BGH implies that PCA is not satisfied while SC is satisfied. This representation of inefficiency always occurs as long as $k > 1$, since a higher innovation cost hinders firm A from investing in the innovation process to achieve IT.

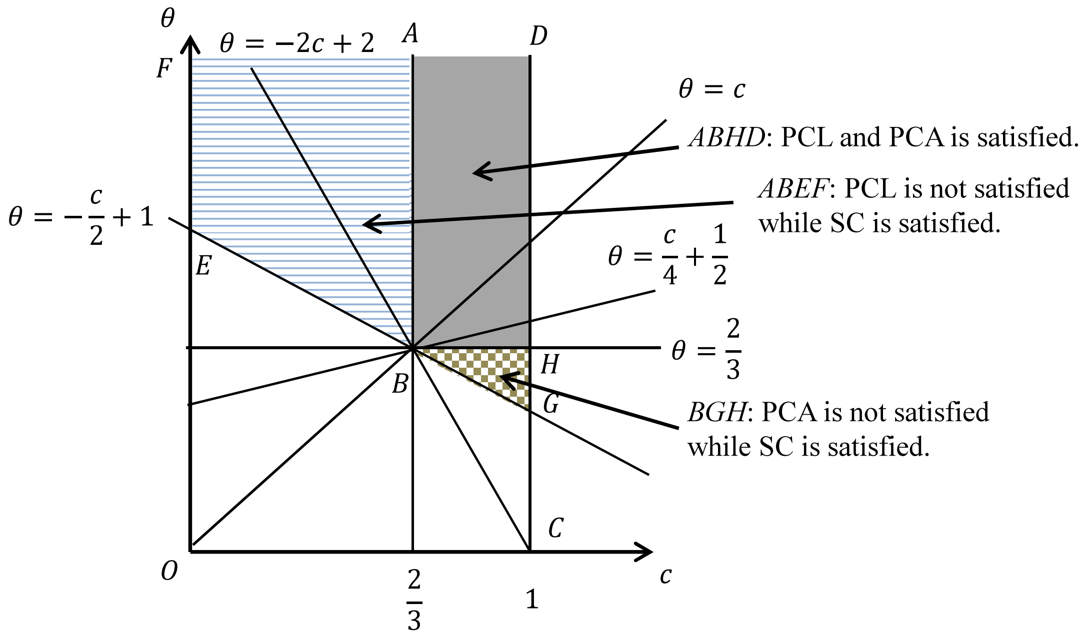


Figure 2.5. Diagram of (c, θ) (severable innovation; $k = 2$).

To summarize, whenever an initial license contract is agreed upon by both firms L and A, socially desirable innovation is always realized. But there is also a possibility that although concluding an initial license contract is socially desirable, either firm L or A does not agree with it, and as a consequence, social welfare is not improved.

■ **Case 4: SA**

The NBS in Stage 1 is exactly the same as the above-mentioned Case 2 (NA). As observed from Equation (2.12), since firm A incorporates an innovation cost, k, into its profit when a grant-back contract is negotiated, firm L should eventually consider bearing the cost along with a grant-back. The result is that a license contract including a grant-back clause is concluded if and only if social welfare is maximized. Hence, there is some room to hope that social inefficiency attached to severable innovation can be resolved with the inclusion of a grant-back clause.

The profits and social welfare from Cases 1 to 4 are exhibited in Table 2.6. The following proposition summarizes the results suggested by a strictly positive innovation cost.

	Case 1: NN	Case 2: NA	Case 3: SN		Case 4: SA
			$\theta > c$	$\theta < c$	
π_L	$2+\theta-\frac{c+k}{2}$	$2+\theta-\frac{c+k}{2}$	$1+\frac{c}{2}$	$1+\frac{\theta}{2}$	$2+\theta-\frac{c+k}{2}$
π_A	$\theta+\frac{c-k}{2}$	$\theta+\frac{c-k}{2}$	$1+2\theta-\frac{c}{2}-k$	$1+\frac{3\theta}{2}-k$	$\theta+\frac{c-k}{2}$
W	$2+2\theta-k$	$2+2\theta-k$	$2+2\theta-k$	$2+2\theta-k$	$2+2\theta-k$

Table 2.6. Profits and social welfare (positive innovation cost).

Proposition 2.3 (Positive innovation cost and a single licensee)

(1) With regard to non-severable innovation, irrespective of the availability of a grant-back clause, innovation is achieved and the improved technology is shared by firms L and A if and only if social welfare can be improved by the innovation; and

(2) With regard to severable innovation, if a grant-back clause is not available, there are some possibilities that either firm L or A may not conclude an initial license contract, and hence, the innovation is not achieved in spite of socially desirable innovation. But this social inefficiency can be resolved through a license contract that includes a grant-back clause.

See again Figure 2.5. The areas, *ABEF* and *BGH*, represent the configuration of (c, θ), where PCL and PCA are not satisfied while SC is satisfied, respectively. Since a contract that includes a grant-back clause enhances social welfare only when it is concluded, these two inefficient areas can be eliminated through such a grant-back.

Additionally, while the area, $OCGE$, does not satisfy PCL or PCA, it is also undesirable in terms of social welfare. Considering the result of Case 4 (SA), we can therefore say that a grant-back clause plays a role in resolving a sort of inconsistency that may exist between the participation conditions (PCL and PCA) and social optimality (SC) condition when the innovation is severable.

It is noteworthy that when an innovation cost is strictly positive, the use of a grant-back clause can improve the incentive of a licensee as well as that of a licensor. We find that while the cost of developing innovation is shared by the two firms for non-severable innovation, it is incurred only by firm A for severable innovation. Simply put, this difference depends on the timing of the bargaining. With regard to severable innovation in the absence of a grant-back clause, the innovation cost has already been initiated when the two firms negotiate the license of IT in Stage 2, but not in Stage 1, where no contracts are assumed to be concluded. That is why firm L does not need to take into account any innovation costs when a decision is made to attach the license of BT in Stage 1. In other words, firm L avoids a sharing of the innovation cost in return for giving up a compensation for transferring BT on ground that the innovation is "severable". Thus, it follows that while firm L is not given an opportunity to impose any royalty payment on firm A for severable innovation, it is not required to include an innovation cost in the initial license contract either. In this regard, firm A needs to inherit the entire innovation cost, not placing any burden on firm L. By contrast, under the scheme of a grant-back clause, firm L is compelled to carefully consider the innovation cost of IT on the premise of receiving IT in negotiation with firm A, while being able to set the royalty rate of BT in Stage 1. In short, firm L is partly responsible for the innovation cost in a grant-back contract in order to ensure firm A's success in follow-on innovation. This inevitably leads to the sharing of the innovation cost between firms L and A, and thereby, contributes to reducing the burden on firm A, which also results in an increase in the profits of both firms L and A due to the achievement of an eventual follow-on innovation.

As observed above, such an *ex ante* contract works as a direct tool to encourage a licensee to innovate as well as a licensor to transfer her base technology, the former of which is precisely the advantage that Scotchmer (1991, 1996) stresses in her study where innovation is costly as it is here.

2.5. Two heterogeneous licensees

When there are more than a single licensee, innovation achieved by one of the licensees jeopardizes not only the licensor's market but also the other licensees' markets. This has a feedback effect on the profits through bargaining over royalty payments between a licensor and licensees. In order to highlight this effect, it is assumed that there are two heterogeneous licensees and that only a single licensee can afford to innovate once she has been granted the license of BT. This "competent" licensee is labeled as firm A and the other "incompetent" licensee as firm B. (The set of the firms is denoted as $F = \{L, A, B\}$.) We have to take note that the royalty rates of BT paid to firm L may vary between firms A and B, so that r_1^A (the royalty rate of firm A) needs not be equivalent to r_1^B (the royalty rate of firm B).

This section consists of many simplified assumptions so as not to make the analyses more complicated than they are necessary. It is therefore postulated that the innovation cost, k, is zero and that territorial restrictions are not prohibited. The purpose of this section here is to mere illustrate some degree of the robustness of our results obtained so far based on simple assumptions, but not to conduct a full analysis. To come right to the point, the “but for" defense of a grant-back clause for severable innovation is still valid even if multiple heterogeneous licensees in terms of innovation abilities are introduced in the model. Section 2.7 reveals that although the absence of territorial restrictions in part slightly affect the results, the nature of the “but for" defense remains essentially the same.

2.5.1. Non-severable innovation

The timing of the game is modified in what follows. In Stage 0, firm L already possesses BT and the attribute of the innovation is known to firms L, A, and B. In Stage 1, firm L offers both firms A and B contracts regarding the license of BT. In Stage 2, firm A offers both firms L and B contracts regarding the acquisition of the license of IT. After these stages, firms L, A, and B sell their products and the profits are realized. Figure 2.6 depicts the timing of this multistage game. As described later, the results are not affected by the timing determined in each stage.

[Stage 0]

Firm L already possesses BT.
The attribute of the innovation is known to firms L, A, and B.

[Stage 1]

Firm L decides:
- to license BT to firm A; and
- to include a grant-back clause in a license contract.

• If a license contract is concluded, innovation is immediately achieved and the game proceeds to Stage 2.
• If a license contract is not concluded, firm L enters the market of firm A.

Firm L decides to license BT to firm B

• If a license contract is not concluded, firm L enters the market of firm B.

[Stage 2]

(If a grant-back clause is not included) Firm A decides to license IT to firm L.

Firm L (or firm A) decides to license IT to firm B.

The profits of firms L, A, and B are realized.

Figure 2.6. Timing of the model (multiple licensees).

An additional assumption is made that firm B, which is eligible to acquire the license of IT from firm A in Stage 2, can have full control of IT only when being granted the license of BT by firm L in Stage 1. In this sense, mastering BT is considered indispensable for firm B to employ IT in its new market, on ground that it is impossible for firm B to apply the result of innovation achieved by firm A without learning BT. Moreover, as assumed before, if firms A and B conclude a contract with firm L regarding the license of BT in Stage 1, their markets are mutually bounded by the territorial restrictions that comprise an integral element of license contracts.

Non-availability of grant-back clause (NN)

Let us suppose that a grant-back clause is not available for non-severable innovation. The cases are separated in accordance with how firm L provides a license contract in Stage 1. There are possible four cases: “firm L does not conclude a contract with firms A and B" (Case 1-NN), “firm L concludes a contract only with firm B" (Case 2-NN), “firm L concludes a contract only with firm A" (Case 3-NN), and “firm L concludes contracts with both firms A and B (Case 4-NN). Table 2.7 summarizes the royalties exchanged between the firms.

r_{1A}	Firm L transfers BT to firm A in Stage 1
r_{1B}	Firm L transfers BT to firm B in Stage 1
r_{2L}	Firm A transfers IT to firm L in Stage 2
r_{2B}	Firm A transfers IT to firm B in Stage 2

Table 2.7. Royalty rates exchanged between the firms.

■ **Case 1-NN**: Firm L does not conclude a contract with firms A and B.

With restrictions in place preventing firm A from innovating in the absence of BT, firms L, A, and B never proceed to Stage 2. Firm L earns $\pi_L^{NN1} = 1 + (1 - c) + (1 - c) = 3 - 2c$ by entering other markets at the expense of the market entry cost, c, each. By contrast, the earnings of firms A and B amount to zero: $\pi_A^{NN1} = \pi_B^{NN1} = 0$. Hence, social welfare is pushed down to the lowest level, $W^{NN1} = \sum_{f \in F} \pi_f^{NN1} = 3 - 2c$.

■ **Case 2-NN**: Firm L concludes a contract only with firm B.

It sounds somewhat odd that firm L does not come to a contract agreement with firm A, but only with firm B, because the latter firm does not have a competence to innovate further. But such a license contract can be interpreted as establishing a “virtual" territorial restriction and saving the market entry cost incurred by firm L. [21] We obtain the NBS between firms L and B in Stage 1 from the following. (Stage 2 is omitted.)

$$\max_{r_{1B}}[(1 + 1 - c + r_{1B}) - (3 - 2c)][(1 - r_{1B}) - 0]$$
$$= \max_{r_{1B}}(-1 + c + r_{1B})(1 - r_{1B}). \Rightarrow r_{1B}^{NN2} = 1 - \frac{c}{2}. \tag{2.15}$$

This NBS implies that the market entry cost is equally shared between firms L and B. The profits of the firms and social welfare are summarized as: $\pi_L^{NN2} = 3 - \frac{3c}{2}$, $\pi_A^{NN2} = 0$, $\pi_B^{NN2} = \frac{c}{2}$, and $W^{NN2} = 3 - c$, respectively.

■ **Case 3-NN**: Firm L concludes a contract only with firm A.

A reasonable assumption can be made that firm A is not allowed to transfer BT to firm B due to possible patent right infringement accruing to only firm L. Since IT is not a significant technology for firm B in the absence of BT, firms A and B cannot conclude a contract regarding IT in Stage 2.

By backward induction, we first consider whether firm A licenses IT to firm L in Stage 2. The NBS in Stage 2 is determined by the following problem:

$$\max_{r_{2L}}[(1 + \theta + r_{1A} - r_{2L} + 1 - c) - (1 + r_{1A} + 1 - c)]$$
$$[(1 + \theta + r_{2L} - r_{1A}) - (1 + \theta - r_{1A})]$$
$$= \max_{r_{2L}}(\theta - r_{2L})r_{2L}. \Rightarrow r_{2L}^{NN3} = \frac{\theta}{2}. \tag{2.16}$$

---

[21] This kind of a territorial restriction is at risk for being deemed to be against a competition policy. Eliminating Case 2-NN from a series of the analysis slightly changes the result that is derived later, but the fundamental implication for the “but for" defense remains the same. See following Appendix in Section 2.7 for more details.

Next, the NBS between firms L and A in Stage 1 is provided by:

$$\max_{r_{1A}}[(1+\theta+r_{1A}-r_{2L}^{NN3}+1-c)-(3-2c)][(1+\theta+r_{2L}^{NN3}-r_{1A})-0]$$
$$=\max_{r_{1A}}\left(-1+\frac{\theta}{2}+c+r_{1A}\right)\left(1+\frac{3\theta}{2}-r_{1A}\right). \Rightarrow r_{1A}^{NN3}=1+\frac{\theta}{2}-\frac{c}{2}. \tag{2.17}$$

From Equations (2.16) and (2.17), the profits of the firms and social welfare are summarized as: $\pi_L^{NN3} = 3+\theta-\frac{3c}{2}$, $\pi_A^{NN3} = \theta+\frac{c}{2}$, $\pi_B^{NN3} = 0$, and $W^{NN3} = 3+2\theta-c$, respectively. Because $\pi_L^{NN3} > \pi_L^{NN2}$ holds, firm L prefers to conclude an initial license contract with firm A rather than with firm B. Also, social welfare increases due to the innovation of IT and its use in the markets of both firms L and A.

■ **Case 4-NN**: Firm L concludes contracts with both firms A and B.

Assuming that firm L has already granted BT to firms A and B in Stage 1, we first focus on the NBS between firms A and B in Stage 2 where they conclude a contract regarding the license of IT. The NBS in Stage 2 is provided by:

$$\max_{r_{2B}}\underbrace{[(1+\theta+r_{2L}+r_{2B}-r_{1A})-(1+\theta+r_{2L}-r_{1A})]}_{\text{Profit difference of firm A}}$$
$$\underbrace{[(1+\theta-r_{1B}-r_{2B})-(1-r_{1B})]}_{\text{Profit difference of firm B}}$$
$$=\max_{r_{2B}} r_{2B}(\theta-r_{2B}). \Rightarrow r_{2B}^{NN4}=\frac{\theta}{2}. \tag{2.18}$$

The terms in Equation (2.18) is concerning firms A and B from left to right. Firm A may be capable of entering the market of firm B by employing IT when they do not conclude a contract, because the territorial restriction to be in place between them has not been established. However, it is of no surprise that firm L never provides a contract that will place it in a disadvantaged position especially when the innovation is non-severable. Hence, we can argue that firm L offers firm A with the license of BT in Stage 1 on the condition that firm A never actively engages in any specific action within the market of firm B. Since the following analysis relies on this assumption, only firm L is a threat to firm B. Likewise, the NBS between firms L and A in Stage 2 is provided by:

$$\begin{aligned}
&\max_{r_{2L}}[(1+\theta+r_{1A}+r_{1B}-r_{2L})-(1+r_{1A}+r_{1B})]\\
&[(1+\theta-r_{1A}+r_{2L}+r_{2B}^{NN4})-(1+\theta-r_{1A}+r_{2B}^{NN4})]\\
&=\max_{r_{2L}}(\theta-r_{2L})r_{2L}. \Rightarrow r_{2L}^{NN4}=\frac{\theta}{2}.
\end{aligned} \tag{2.19}$$

In Stage 1, the NBS between firms L and B can be obtained as follows:

$$\begin{aligned}
&\max_{r_{1B}}[(1+\theta+r_{1A}-r_{2L}^{NN4}+r_{1B})-(1+\theta+r_{1A}-r_{2L}^{NN4}+1-c)]\\
&[(1+\theta-r_{1B}-r_{2B}^{NN4})-0]\\
&=\max_{r_{1B}}(-1+c+r_{1B})\left(1+\frac{\theta}{2}-r_{1B}\right). \Rightarrow r_{1B}^{NN4}=1+\frac{\theta}{4}-\frac{c}{2}.
\end{aligned} \tag{2.20}$$

Lastly, the NBS between firms L and A is considered. The important point is that the reservation profit of firm L must be $1+(1-c)+(1-\frac{c}{2}) = 3-\frac{3c}{2}$ (but not $3-2c$) since firm L simultaneously can offer firm B with a contract that could constitute a virtual territorial restriction. Consequently, the NBS between firms L and A in Stage 1 is determined by:

$$\max_{r_{1A}}\left[(1+\theta+r_{1A}-r_{2L}^{NN4}+r_{1B}^{NN4})-\left(3-\frac{3c}{2}\right)\right][(1+\theta+r_{2L}^{NN4}-r_{1A}+r_{2B}^{NN4})-0]$$

$$= \max_{r_{1A}} \left(-1 + \frac{3\theta}{4} + c + r_{1A}\right)(1 + 2\theta - r_{1A}). \Rightarrow r_{1A}^{NN4} = 1 + \frac{5\theta}{8} - \frac{c}{2}. \tag{2.21}$$

Hence, the profits of the firms and social welfare are summarized as: $\pi_L^{NN4} = 3 + \frac{11\theta}{8} - c$, $\pi_A^{NN4} = \frac{11\theta}{8} + \frac{c}{2}$, $\pi_B^{NN4} = \frac{\theta}{4} + \frac{c}{2}$, and $W^{NN3} = 3 + 3\theta$, respectively.

Table 2.8 represents the profits of the firms and social welfare from Cases 1-NN to 4-NN. Comparing these cases, it can be easily determined that preference would be given to Case 4-NN by firms L, A, and B.

	Case 1-NN	Case 2-NN	Case 3-NN	Case 4-NN
π_L	$3 - 2c$	$3 - \frac{3c}{2}$	$3 + \theta - \frac{3c}{2}$	$3 + \frac{11\theta}{8} - c$
π_A	0	0	$\theta + \frac{c}{2}$	$\frac{11\theta}{8} + \frac{c}{2}$
π_B	0	$\frac{c}{2}$	0	$\frac{\theta}{4} + \frac{c}{2}$
W	$3 - 2c$	$3 - c$	$3 + 2\theta - c$	$3 + 3\theta$

Table 2.8. Profits and social welfare (NN).

Lemma 2.4 (Non-severable innovation and two heterogeneous licensees) When a grant-back clause is not included in a license contract:

(1) Innovation is achieved and the improved technology is shared by all firms L, A, and B through the license contracts concluded between them;

(2) First-best social welfare is generated; and

(3) The distribution of profits is Pareto optimal.

Proof (1) $\pi_L^{NN4} - \pi_L^{NN3} = (3 + \frac{11\theta}{8} - c) - (3 + \theta - \frac{3c}{2}) = \frac{3\theta}{8} + \frac{c}{2} > 0 \Leftrightarrow \pi_L^{NN4} > \pi_L^{NN3}$. This means that firm L prefers Case 4-NN to Case 3-NN. It is also clear that $\pi_L^{NN4} > \pi_L^{NN3} > \pi_L^{NN2} > \pi_L^{NN1}$. It can be easily shown that $\pi_A^{NN4} = \frac{11\theta}{8} + \frac{c}{2} > \pi_A^{NN2} = \pi_A^{NN1} = 0$ (firm A does not conclude a contract with firm L) and $\pi_B^{NN4} = \frac{\theta}{4} + \frac{c}{2} > \pi_B^{NN3} = \pi_B^{NN1} = 0$ (firm

B does not conclude a contract with firm L). Hence, both firms A and B also prefer to accept the series of the contracts offered by firm L in Case 4-NN.

(2)(3) It is clear that we can obtain $W^{NN4} > W^{NN3} > W^{NN2} > W^{NN1}$, $\pi_L^{NN4} > \pi_L^{NN3} > \pi_L^{NN2} > \pi_L^{NN1}$, $\pi_A^{NN4} > \pi_A^{NN3} > \pi_A^{NN2} = \pi_A^{NN1}$, and $\pi_B^{NN4} > \pi_B^{NN2} > \pi_B^{NN3} = \pi_B^{NN1}$. While Case 4-NN generates first-best social welfare, the distribution is certain to maintain Pareto optimal. ■

The net royalties of firms L, A, and B are represented by $r_L^{NN4} = r_{1A}^{NN4} - r_{2L}^{NN4} + r_{1B}^{NN4} = 2 + \frac{3\theta}{8} - c$, $r_A^{NN4} = r_{2L}^{NN4} - r_{1A}^{NN4} + r_{2B}^{NN4} = -1 + \frac{3\theta}{8} + \frac{c}{2}$, and $r_B^{NN4} = -r_{1B}^{NN4} - r_{2B}^{NN4} = -1 - \frac{3\theta}{4} + \frac{c}{2}$, respectively. Obviously, the sum of these net royalties always amounts to zero (i.e. $r_L^{NN3} + r_A^{NN3} + r_B^{NN3} = 0$), [22] and bargaining over the royalty payments does not exacerbate any social value. We need to notice that the underlying mechanism is basically the same as in the case of a single licensee for non-severable innovation. Anticipating that follow-on innovation creates benefits in new markets of firms A and B, firm L can in advance capture them through imposing the royalties of the initial license contracts that are offered for both firms A and B in Stage 1. Such use of royalties stems from the assumption that the markets are kept separate due to territorial restrictions. Firm L is, therefore, likely to provide the license of BT to all the licensees, which leads to the follow-on innovation and its sharing among all the firms.

[22] Whereas $r_L^{NN4} > 0$ and $r_B^{NN4} < 0$ always hold, the sign of r_A^{NN4} is indecisive. The condition for $r_A^{NN4} > 0$ is provided by $\theta > \frac{4}{3}(2 - c)$.

Availability of a grant-back clause (NA)

In the case where a grant-back clause is used, firm A can no longer freely provide firm B with IT due to the patent right that belongs to firm L. There are two directions that firm L can take if it intends to disseminate IT to firm B. The first is that firm L directly offers a contract regarding the license of IT to firm B in Stage 2 without involving other parties rigorously based on the nature of a grant-back clause (that is, the patent right of IT belongs solely to firm L). The second is that firm L delegates firm A with the task of transferring IT to firm B by including such an obligation in the initial license contract. In other words, the use of IT is conditional on disseminating it to other firms as well as on granting back to the original licensor. In what follows, we demonstrate that it is profitable for firm L to opt for the second direction.

In accordance with the first direction, the timing of the game is slightly modified as follows: in Stage 1, firm L offers firms A and B a contract regarding the license of BT, and in Stage 2, firm L offers firm B with a contract regarding the license of IT. In a similar fashion, the cases should be classified. The discussions of Cases 1-NA and 2-NA, which are equivalent to the previous analyses, are omitted. Instead, Case 5-NA (Firm L concludes contracts with firms A and B in Stage 1, but not with firm B in Stage 2) should be added.

■ **Case 3-NA**: Firm L concludes a contract only with firm A.

This case implies that while firm A achieves innovation in Stage 1, firm L is reluctant to share both BT and IT with firm B. The NBS between firms L and A in Stage 1 is provided by:

$$\max_{r_{1A}}[(1+\theta+r_{1A}+1-c)-(3-2c)][(1+\theta-r_{1A})-0]$$
$$=\max_{r_{1A}}(-1+\theta+c+r_{1A})(1+\theta-r_{1A}). \Rightarrow r_{1A}^{NA3}=1-\frac{c}{2}. \qquad (2.22)$$

The profits of the firms and social welfare are summarized as: $\pi_L^{NA3} = 3+\theta-\frac{3c}{2}$, $\pi_A^{NA3}=\theta+\frac{c}{2}$, $\pi_B^{NA3}=0$, and $W^{NA3}=3+2\theta-c$, respectively. We obtain $\pi_L^{NA3}=\pi_L^{NN3}$, because firm L can charge the same net royalty rate in the absence of a grant-back clause as it is equal to the royalty rate induced by a grant-back clause for non-severable innovation.

■ **Case 4-NA**: Firm L concludes contracts with both firms A and B.

Option 1: Firm L offers a contract regarding the license of IT directly to firm B.

By assuming that firm L has already concluded contracts with firms A and B regarding the license of BT in Stage 1, we consider the NBS between firms L and B in Stage 2:

$$\max_{r_{2B}}[(1+\theta+r_{1A}+r_{1B}+r_{2B})-(1+\theta+r_{1A}+r_{1B})]$$
$$[(1+\theta-r_{1B}-r_{2B})-(1-r_{1B})]$$
$$=\max_{r_{2B}} r_{2B}(\theta-r_{2B}). \Rightarrow r_{2B}^{NA4(1)}=\frac{\theta}{2}. \qquad (2.23)$$

We also consider the NBS between firms L and B in Stage 1: [23]

$$\max_{r_{1A}}[(1+\theta+r_{1A}+r_{1B}+r_{2B}^{NA4(1)})-(1+\theta+r_{1A}+1-c)]$$
$$[(1+\theta-r_{1B}-r_{2B}^{NA4(1)})-0]$$

[23] The following result does not depend on the timing of the contracts in Stage 1.

$$= \max_{r_{1A}} \left(-1 + \frac{\theta}{2} + c + r_{1B}\right)\left(1 + \frac{\theta}{2} - r_{1B}\right). \Rightarrow r_{1B}^{NA4(1)} = 1 - \frac{c}{2}. \tag{2.24}$$

Finally, the NBS between firms L and A in Stage 1 is provided by:

$$\begin{aligned}&\max_{r_{1A}}[(1 + \theta + r_{1A} + r_{1B}^{NA4(1)} + r_{2B}^{NA4(1)}) - (1 + 1 - c + r_{1B}^{NA4(1)})][(1 + \theta - r_{1A}) - 0]\\ &= \max_{r_{1A}} \left(-1 + \frac{3\theta}{2} + c + r_{1A}\right)(1 + \theta - r_{1A}). \Rightarrow r_{1A}^{NA4(1)} = 1 - \frac{\theta}{4} - \frac{c}{2}.\end{aligned} \tag{2.25}$$

$r_{1B}^{NA4(1)} = 1 - \frac{c}{2}$ (but not $1 - c$) is the potential profit of firm L related to the handling of negotiations with firm B. From Equation (2.25), we obtain $r_{1A}^{NA4(1)} = 1 - \frac{\theta}{4} - \frac{c}{2}$.[24] Hence, the profits of the firms and social welfare are summarized as: $\pi_L^{NA4(1)} = 3 + \frac{5\theta}{4} - c$, $\pi_A^{NA4(1)} = \frac{5\theta}{4} + \frac{c}{2}$, $\pi_B^{NA4(1)} = \frac{\theta}{2} + \frac{c}{2}$, and $W^{NA4(1)} = 3 + 3\theta$, respectively. Since both BT and IT are transferred to every firm, social welfare amounts to the first-best level.

Option 2: Firm L delegates firm A with the task of transferring IT to firm B.

Given that firms A and B are licensed BT in Stage 1, we consider the NBS regarding the license of IT between the two firms in Stage 2 as follows:

$$\begin{aligned}&\max_{r_{2B}}[(1 + \theta + r_{2B}) - (1 + \theta)][(1 + \theta - r_{1B} - r_{2B}) - (1 - r_{1B})]\\ &= \max_{r_{2B}} r_{2B}(\theta - r_{2B}). \Rightarrow r_{2B}^{NA4(2)} = \frac{\theta}{2}.\end{aligned} \tag{2.26}$$

Let us revert back to Stage 1 where firm L transfers BT to firms A and B. In the first place, we derive the NBS between firms L and B in Stage 1 as follows:

$$\max_{r_{1A}}[(1 + \theta + r_{1A} + r_{1B}) - (1 + \theta + r_{1A} + 1 - c)][(1 + \theta - r_{1B} - r_{2B}^{NA4(2)}) - 0]$$

[24] The condition for $r_{1A}^{NA4} > 0$ is provided by $\theta < 2(2 - c)$. If this condition is not satisfied, the royalty payment flows from firm L to firm A.

$$= \max_{r_{1A}}(-1 + c + r_{1B})\left(1 + \frac{\theta}{2} - r_{1B}\right). \Rightarrow r_{1B}^{NA4(2)} = 1 + \frac{\theta}{4} - \frac{c}{2}. \tag{2.27}$$

Next, the NBS between firms L and A is provided by:

$$\max_{r_{1A}} \left[(1 + \theta + r_{1A} + r_{1B}^{NA4(2)}) - \left(1 + 1 - c + 1 - \frac{c}{2}\right)\right] [(1 + \theta - r_{1A} + r_{2B}^{NA4(2)}) - 0]$$
$$= \max_{r_{1A}} \left(-1 + \frac{5\theta}{4} + c + r_{1A}\right)\left(1 + \frac{3\theta}{2} - r_{1A}\right). \Rightarrow r_{1A}^{NA4(2)} = 1 + \frac{\theta}{8} - \frac{c}{2}. \tag{2.28}$$

From these, the profits of the firms and social welfare are summarized as: $\pi_L^{NA4(2)} = 3 + \frac{11\theta}{8} - c$, $\pi_A^{NA4(2)} = \frac{11\theta}{8} + \frac{c}{2}$, $\pi_B^{NA4(2)} = \frac{\theta}{4} + \frac{c}{2}$, and $W^{NA4(2)} = 3 + 3\theta$. Clearly, the result is totally the same as that of Case 4-NN. The reason is that as described before, firm L can deliberately set royalty rates so as to secure benefits in the timing of the initial contracts. By comparing the profit of firm L, it can be demonstrated that firm L necessarily prefers to choose Option 2.

Lemma 2.5 Suppose that firm L can require firm A to transfer IT to firm B by a grant-back contract in return for the use of IT. Then, firm L prefers to conclude such a contract rather than directly transfers IT to firm B by itself.

Proof Since $\pi_L^{NA4(2)} - \pi_L^{NA4(1)} = (3 + \frac{11\theta}{8} - c) - (3 + \frac{5\theta}{4} - c) = \frac{\theta}{8} > 0$, we obtain $\pi_L^{NA4(2)} > \pi_L^{NA4(1)}$. Hence, firm L prefers to choose Option 2. ■

Intuitively, since firm L can save a bargaining cost concerning the contract regarding the license of IT with firm B through delegating this task to firm A, it can earn a higher profit. Put it differently, when firm L needs to directly incur a bargaining cost by negotiating with firm B to license IT, the profit of firm L is likely to fall by contrary. It is noticeable that firm A is also likely to increase its profit by this grant-back contract because it can receive a royalty payment from firm B in return for its IT. [25]

For descriptive purposes, using the result of Option 2, we can represent the profits and social welfare as follows: $\pi_L^{NA4} = 3 + \frac{11\theta}{8} - c$, $\pi_A^{NA4} = \frac{11\theta}{8} + \frac{c}{2}$, $\pi_B^{NA4} = \frac{\theta}{4} + \frac{c}{2}$, and $W^{NA4} = 3 + 3\theta$.

■ **Case 5-NA**: Firm L concludes contracts with firms A and B in Stage 1, but not with firm B in Stage 2.

In this case, firm L intends to establish a territorial restriction with firm B to save a market entry cost, but is reluctant to make firm A share the outcome of innovation with firm B. First, the NBS between firms L and B in Stage 1 is provided by: [26]

$$\begin{aligned}&\max_{r_{1B}}[(1+\theta+r_{1A}+r_{1B})-(1+\theta+r_{1A}+1-c)][(1-r_{1B})-0]\\&=\max_{r_{1B}}(-1+c+r_{1B})(1-r_{1B}).\Rightarrow r_{1B}^{NA5}=1-\frac{c}{2}.\end{aligned}\tag{2.29}$$

Next, the NBS between firms L and A in Stage 1 is provided by:

$$\begin{aligned}&\max_{r_{1A}}[(1+\theta+r_{1A}+r_{1B}^{NA5})-(1+1-c+r_{1B}^{NA5})][(1+\theta-r_{1A})-0]\\&=\max_{r_{1A}}(-1+\theta+c+r_{1A})(1+\theta-r_{1A}).\Rightarrow r_{1A}^{NA5}=1-\frac{c}{2}.\end{aligned}\tag{2.30}$$

From Equations (2.29) and (2.30), the profits of the firms and social welfare are summarized as: $\pi_L^{NA5} = 3+\theta-c$, $\pi_A^{NA5} = \theta+\frac{c}{2}$, $\pi_B^{NA5} = \frac{c}{2}$, and $W^{NA5} = 3+2\theta$, respectively.

---

[25] Since $\pi_A^{NA4(2)} - \pi_A^{NA4(1)} = (\frac{11\theta}{8} + \frac{c}{2}) - (\frac{5\theta}{4} + \frac{c}{2}) = \frac{\theta}{8} > 0$, we obtain $\pi_A^{NA4(2)} > \pi_A^{NA4(1)}$.

[26] As observed from Equations (2.29) and (2.30), the following result does not depend on the timing of the contracts in Stage 1.

Proposition 2.4 (Non-severable innovation and two heterogeneous licensees) Suppose that a grant-back clause as assumed in Lemma 2.5 is available in a license contract. Then, we can achieve the same results as indicated in Lemma 2.4 by using a contract that includes the grant-back clause. Hence, the grant-back clause has no effects on profits or on social welfare.

Proof Firm L prefers Case 4-NA rather than Case 5-NA when a grant-back clause is available, since $\pi_L^{NA4} - \pi_L^{NA5} = (3 + \frac{11\theta}{8} - c) - (3 + \theta - c) = \frac{3\theta}{8} > 0 \Leftrightarrow \pi_L^{NA4} > \pi_L^{NA5}$. Similarly, it can be easily demonstrated that Case 4-NA generates the highest profit for firm L among the five cases. Comparing the maximum profits of firms and social welfare between when a grant-back clause is made available and when it is not, we obtain $\pi_L^{NA4} = \pi_L^{NN4}$, $\pi_A^{NA4} = \pi_A^{NN4}$, $\pi_B^{NA4} = \pi_B^{NN4}$, and $W^{NA4} = W^{NN4}$. Consequently, Lemma 2.4 (1) to (3) directly apply. ■

Table 2.9 presents the profits of the firms and social welfare when a grant-back clause that stipulates an obligation of transferring IT to firm B is available. Proposition 2.4 demonstrates that when there are heterogeneous licensees for non-severable innovation, the availability of a grant-back clause does not affect firms' profits and social welfare. This argument is exactly identical to Lemma 2.1 which maintains that when there is a single competent licensee for non-severable innovation, a grant-back clause does not affect any results. In other words, the “equivalence" of making use of a grant-back clause applies to both a single competent licensee and multiple heterogeneous licensees.

	Case 1-NA	Case 2-NA	Case 3-NA	Case 4-NA	Case 5-NA
π_L	$3-2c$	$3-\frac{3c}{2}$	$3+\theta-\frac{3c}{2}$	$3+\frac{11\theta}{8}-c$	$3+\theta-c$
π_A	0	0	$\theta+\frac{c}{2}$	$\frac{11\theta}{8}+\frac{c}{2}$	$\theta+\frac{c}{2}$
π_B	0	$\frac{c}{2}$	0	$\frac{\theta}{4}+\frac{c}{2}$	$\frac{c}{2}$
W	$3-2c$	$3-c$	$3+2\theta-c$	$3+3\theta$	$3+2\theta$

Table 2.9. Profits and social welfare (NA).

This result depends on the assumption that a grant-back clause is modified to include the new obligation indicated in the above. Otherwise, firm L chooses to make an initial license contract that does not include a grant-back clause even if it is made available. [27] Although one may doubt the true feasibility of such a grant-back clause,, firms L and A actually benefit from it as was demonstrated. Because of this, it stands to reason for them to agree on allowing firm L to delegate firm A with the task of transferring IT to firm B. Hence, we can point out that the grant-back clause, which allows a competent licensee to freely use and license her improved technology to other firms, do benefit social welfare.

2.5.2. Severable innovation

Non-availability of a grant-back clause

We classify the cases based on whether the value of innovation is larger or smaller than the market entry cost. The discussions of Cases 1-SN and 2-SN are omitted to save space because they are equivalent to Cases 1-NN and 2-NN discussed earlier.

[27] Let us focus on the profit of Case 4-NA in Option 1. We can then derive $\pi_L^{NA4(1)} - \pi_L^{NN4} = (3+\frac{5\theta}{4}-c)-(3+\frac{11\theta}{8}-c) = -\frac{\theta}{8} < 0 \Leftrightarrow \pi_L^{NA4(1)} < \pi_L^{NN4}$. In this environment, firm L will therefore choose an initial license contract that does not include a grant-back clause even if it is available.

■ **Case 3-SN**: Firm L concludes a contract only with firm A.

Both firms L and A are determined to enter the original and new markets of firm B because no territorial restrictions whatsoever have been established with firm B. On this point, an assumption is somehow required regarding the consequence of competition in the markets of firm B. One probable assumption could be that firm L obtains the original market (value of 1) while firm A does the new market created by the innovation (value of θ). This division of the whole market of firm B is exactly a mirror to the competition in the markets of firm L, where both firms L and A possess IT. That is, the way firms L and A divide the original and new markets of firm B is determined by the market power in the original market (firm L obtains 1) and new market (firm A obtains θ) of firm L under severable innovation. Although this assumption seems a bit arbitrary to readers, it is the most reasonable among others. Therefore, we posit that since this market power precisely reflects the relationship between firms L and A, firm L retains the original market and firm A gains entry to the new market. [28]

When $\theta > c$, the NBS between firms L and A in Stage 2 is provided by:

$$\max_{r_{2L}}[(1+\theta-r_{2L}+1-c)-(2-c)][(1+\theta+r_{2L}+\theta-c)-(1+\theta+\theta-c+\theta-c)]$$

$$=\max_{r_{2L}}(\theta-r_{2L})(-\theta+c+r_{2L}). \Rightarrow r_{2L}^{SN3}=\theta-\frac{c}{2}. \tag{2.31}$$

Because the innovation is severable, firm L is not in a position to demand any royalty payments in return for BT from firm A in Stage 1. The profits of the firms and social welfare are summarized as: $\pi_L^{SN3}=2-\frac{c}{2}$, $\pi_A^{SN3}=1+3\theta-\frac{3c}{2}$, $\pi_B^{SN3}=0$, and $W^{SN3}=$

[28] Since innovation is severable, it is plausible that firm A has been granted the right to freely use IT, and accordingly, has every intention of seeking entry to the new market of firm B using IT.

$3 + 3\theta - 2c$, respectively.

When $\theta < c$, firm A cannot foresee any advantage of entering the new markets of both firms L and B. The NBS between firms L and A is provided by:

$$\begin{aligned}&\max_{r_{2L}}[(1+\theta-r_{2L}+1-c)-(2-c)][(1+\theta+r_{2L})-(1+\theta)]\\&=\max_{r_{2L}}(\theta-r_{2L})r_{2L}. \Rightarrow r_{2L}^{SN2}=\frac{\theta}{2}.\end{aligned} \tag{2.32}$$

The profits of the firms and social welfare are summarized as: $\pi_L^{SN3} = 2 + \frac{\theta}{2} - c$, $\pi_A^{SN3} = 1 + \frac{3\theta}{2}$, $\pi_B^{SN3} = 0$, and $W^{SN3} = 3 + 2\theta - c$, respectively.

With regard to social welfare, although innovation is actually achieved by firm A for any (θ, c), the market entry cost, c or $2c$, must be deducted from social welfare.

■ **Case 4-SN**: Firm L concludes contracts with both firms A and B.

It is plausible to assume that firm L cannot employ IT that has been licensed by firm A in the new market of firm B for severable innovation. A presupposition could be that firm A does not allow firm L to use IT anywhere else other than its own new market by laying claim to its patent right when bargaining over the license of IT.

When $\theta > c$, the NBS between firms A and B in Stage 2 is provided by:

$$\begin{aligned}&\max_{r_{2B}}[(1+\theta+r_{2L}+r_{2B})-(1+\theta+r_{2L}+\theta-c)][(1+\theta-r_{2B})-1]\\&=\max_{r2B}(-\theta+c+r_{2B})(\theta-r_{2B}). \Rightarrow r_{2B}^{SN4}=\theta-\frac{c}{2}.\end{aligned} \tag{2.33}$$

It is necessary to bear in mind that the use of IT by firm B in Stage 2 does not generate any royalty payments associated with BT in severable innovation. In addition, we derive the NBS between firms L and A in Stage 2 for $\theta > c$ as follows:

$$\max_{r_{2L}}[(1+\theta-r_{2L})-1]][(1+\theta+r_{2L}+r_{2B}^{SN4})-(1+\theta+\theta-c+r_{2B}^{SN4})]$$
$$=\max_{r_{2L}}(\theta-r_{2L})(-\theta+c+r_{2L}).\Rightarrow r_{2L}^{SN4}=\theta-\frac{c}{2}. \tag{2.34}$$

From Equations (2.33) and (2.34), the profits and social welfare are summarized as: $\pi_L^{SN4} = 1 + \frac{c}{2}$, $\pi_A^{SN4} = 1 + 3\theta - c$, $\pi_B^{SN4} = 1 + \frac{c}{2}$, and $W^{SN4} = 3 + 3\theta$, respectively. Notably, whereas firm A relishes the value of innovation in every market with an additional market entry cost, c, the earning of firm L is generated only from its own original market.

When $\theta < c$, the NBS between firms A and B in Stage 2 is provided by:

$$\max_{r_{2B}}[(1+\theta+r_{2B}+r_{2L})-(1+\theta+r_{2L})][(1+\theta-r_{2B})-1]$$
$$=\max_{r_{2B}} r_{2B}(\theta-r_{2B}).\Rightarrow r_{2B}^{SN4}=\frac{\theta}{2}. \tag{2.35}$$

Lastly, we derive the NBS between firms L and A in Stage 2 for $\theta < c$ as follows:

$$\max_{r_{2L}}[(1+\theta-r_{2L})-1][(1+\theta+r_{2L}+r_{2B}^{SN4})-(1+\theta+r_{2B}^{SN4})]$$
$$=\max_{r_{2L}}(\theta-r_{2L})r_{2L}.\Rightarrow r_{2L}^{SN4}=\frac{\theta}{2}. \tag{2.36}$$

From Equations (2.35) and (2.36), the profits and social welfare are summarized as: $\pi_L^{SN4} = 1 + \frac{\theta}{2}$, $\pi_A^{SN4} = 1 + 2\theta$, $\pi_B^{SN4} = 1 + \frac{\theta}{2}$, and $W^{SN4} = 3 + 3\theta$, respectively. Contrary to the case of $\theta > c$, firm L earns profits both in its original and new markets, and in particular, half the value of innovation comes from the latter market. On the other hand, firm A can reap rewards in each market, obtaining its whole markets $(1+\theta)$ and one half of the new market $(\frac{\theta}{2})$ of each firm L and B. The profits and social welfare from Cases 1-SN to 4-SN are presented in Table 2.10.

	Case 1-SN	Case 2-SN	Case 3-SN		Case 4-SN	
			$\theta > c$	$\theta < c$	$\theta > c$	$\theta < c$
π_L	$3-2c$	$3-\frac{3c}{2}$	$2-\frac{c}{2}$	$2+\frac{\theta}{2}-c$	$1+\frac{c}{2}$	$1+\frac{\theta}{2}$
π_A	0	0	$1+3\theta-\frac{3c}{2}$	$1+\frac{3\theta}{2}$	$1+3\theta-c$	$1+2\theta$
π_B	0	$\frac{c}{2}$	0	0	$1+\frac{c}{2}$	$1+\frac{\theta}{2}$
W	$3-2c$	$3-c$	$3+3\theta-2c$	$3+2\theta-c$	$3+3\theta$	$3+3\theta$

Table 2.10. Profits and social welfare (SN).

Lemma 2.6 (Severable innovation and two heterogeneous licensees) Suppose that a grant-back clause is not available:

(1) An initial license contract between firms L and A cannot be concluded and innovation is not achieved; and

(2) First-best social welfare is never generated.

Proof (1) In the first place, it should be demonstrated that $\pi_L^{SN2} > \pi_L^{SN3}$. With regard to $\theta > c$, $\pi_L^{SN2} - \pi_L^{SN3} = (3-\frac{3c}{2}) - (2-\frac{c}{2}) = 1-c > 0 \Leftrightarrow \pi_L^{SN2} > \pi_L^{SN3}$. In addition, with regard to $\theta < c$, $\pi_L^{SN2} - \pi_L^{SN3} = (3-\frac{3c}{2}) - (2+\frac{\theta}{2}-c) = 1-\frac{\theta}{2}-\frac{c}{2} > 1-c > 0 \Leftrightarrow \pi_L^{SN2} > \pi_L^{SN3}$. Subsequently, we make a comparison between π_L^{SN3} and π_L^{SN4}. With regard to $\theta > c$, $\pi_L^{SN3} - \pi_L^{SN4} = (2-\frac{c}{2}) - (1+\frac{c}{2}) = 1-c > 0 \Leftrightarrow \pi_L^{SN3} > \pi_L^{SN4}$. In addition, with regard to $\theta < c$, $\pi_L^{SN3} - \pi_L^{SN4} = (2+\frac{\theta}{2}-c) - (1+\frac{\theta}{2}) = 1-c > 0 \Leftrightarrow \pi_L^{SN3} > \pi_L^{SN4}$. From these relations, we can obtain $\pi_L^{SN2} > \pi_L^{SN3} > \pi_L^{SN4}$ for every (c, θ). We can also easily show that $\pi_L^{SN2} > \pi_L^{SN1}$. Hence, firm L prefers Case 2-SN to all other cases presented.

(2) $W^{SN4} = 3 + 3\theta$ is the first-best social welfare. However, $W^{SN2} < W^{SN3} < W^{SN4}$ for every (c, θ). ■

Lemma 2.6 suggests that regardless of the relationship between the value of innovation and the market entry cost, firm L is unlikely to conclude an initial license contract with firm A. More precisely, firm L prefers to institute only a virtual territorial restriction with firm B (Case 2-SN) rather than to provide the license of BT to firm A that can lead to further innovation (Case 3-SN). The reason is intuitively as follows. Even if the value of innovation is larger than the market entry cost ($\theta > c$), firm L fails to earn any profit from its own new market due to entry of firm A under severable innovation. By contrast, in view of the relatively smaller value of innovation ($\theta < c$), the innovation value falls below the potential profit for firm L obtained from forming a territorial restriction with firm B. It can be therefore concluded that no innovation is undertaken by firm A due to not being licensed by firm L, and as a consequence, social welfare remains extremely low.

Availability of a grant-back clause (SA)

If a grant-back clause defined in Subsection 2.5.1 such that firm L delegates firm A to transfer IT to firm B is available to firm L, the situation concerning severable innovation can be improved. Remember that the profits of the firms obtained from using a grant-back are unchanged from the case of non-severable innovation. Thus, Table 2.11 is essentially identical to Table 2.9. The following proposition signifies that a grant-back clause certainly improves social welfare for severable innovation.

	Case 1-SA	Case 2-SA	Case 3-SA	Case 4-SA	Case 5-SA
π_L	$3-2c$	$3-\frac{3c}{2}$	$3+\theta-\frac{3c}{2}$	$3+\frac{11\theta}{8}-c$	$3+\theta-c$
π_A	0	0	$\theta+\frac{c}{2}$	$\frac{11\theta}{8}+\frac{c}{2}$	$\theta+\frac{c}{2}$
π_B	0	$\frac{c}{2}$	0	$\frac{\theta}{4}+\frac{c}{2}$	$\frac{c}{2}$
W	$3-2c$	$3-c$	$3+2\theta-c$	$3+3\theta$	$3+2\theta$

Table 2.11. Profits and social welfare (SA).

Proposition 2.5 (Severable innovation and two heterogeneous licensees) If a license contract with the inclusion of a grant-back clause between firms L and A is used, then:
(1) Innovation is achieved and the improved technology is shared by all firms L, A, and B through the license contracts concluded between them. In particular, firm B is also allowed to access the improved technology possessed by firm L; and
(2) First-best social welfare is generated.

Proof (1)(2) We can derive $\pi_L^{SA4}-\pi_L^{SN2}=(3+\frac{5\theta}{4}-c)-(3-\frac{3c}{2})=\frac{11\theta}{8}+\frac{c}{2}>0 \Leftrightarrow \pi_L^{SA4}>\pi_L^{SN2}$. Hence, since firm L has total control over the contract, it concludes a grant-back contract. In addition, it is clear-cut that innovation is achieved, the improved technology is shared by all firms L, A, and B, and first-best social welfare, $W^{SA4}=3+3\theta$, is generated. ■

The license contract that includes a grant-back clause restores a licensor's incentive not only to encourage a competent licensee to innovate, but also to share the improved technology with all firms including an incompetent licensee. In this sense, grant-back clauses are socially desirable for severable innovation, while it is generally anticipated that innovation ceases to unfold without them. In a nutshell, we can still claim the “but for" defense of a grant-back clause for severable innovation in a convincing way even when there are heterogeneous multiple licensees. [29]

2.6. Concluding remarks

This chapter investigated the effect of a grant-back clause in a license contract in accordance with the attributes of innovation that can be divided into “severable" (i.e. non-infringing) and “non-severable" (i.e. infringing) innovation. Typically, since grant-back clauses require a licensee to “give back" further innovation to a licensor without compensation or compensation that is not linked to the value of follow-on innovation, competition authorities are inclined to be concerned that the clauses may lower the licensee’s incentive to innovate further. Indeed, the EU (2004) attempted to make a clear-cut distinction between severable and non-severable innovation. More precisely, the EU (2004) indicated that while a grant-back clause applying to non-severable innovation is considered to be innocuous, the one applying to severable innovation should be treated with much more skepticism. However, this chapter showed that this rule is subject to debate.

With this in mind, by assuming a single licensee under territorial restrictions at the outset, this chapter demonstrated that a grant-back clause does not further increase the original licensor’s incentive to license her base technology for non-severable innovation. By contrast, it was revealed that an increase of her incentive to license in the presence of a grant-back clause is evident for severable innovation, and as a result,

[29] As a robustness check, Section 2.7 attempts to reconfirm the “but for" defence for severable innovation when territorial restrictions are partially prohibited.

follow-on innovation which enhances social welfare can be followed. This finding suggested that the “but for" defense with a particular focus on a licensor is valid for severable innovation, but not for non-severable innovation, which is a polar opposite to the claim made by the EU (2004). Therefore, as regards to contribution to related literature, while the result is different from a general viewpoint that is in favor of the grant-back clause effect on a licensee’s incentive for non-severable innovation, which is typically represented by the EU (2004), this chapter provides a convincing rationale for the “but for" defense of a licensor’s incentive to make severable innovation.

The model was extended on three points. First, in the case where territorial restrictions are prohibited, first-best social welfare is not generated either for sevrable or non-severable innovation. (In particular, innovation is not achieved in the case of severable innovation.) It was demonstrated that social welfare is recovered if a licensor uses a grant-back clause in an initial license contract for the both attributes of innovation. Second, it was observed that a positive innovation cost deteriorates the licensee’s incentive to innovate in addition to the licensor’s for severable innovation. Interestingly enough, it was shown that a grant-back clause enables both a licensor and a licensee to regain their incentives to achieve follow-on innovation by facilitating *ex ante* bargaining of the improved technology between them. Finally, by including two heterogeneous licensees in terms of innovation abilities, it was proven that a grant-back clause can improve the licensor’s incentive to encourage follow-on innovation and to share it with all the licensees. All things considered, we can say that the “but for " defense of a grant-back clause for severable innovation is by and large robust to the variations of the model.

This chapter has consistently cast doubt as to whether the seemingly plausible reasoning presented by the EU (2004) can be used to deny the use of a grant-back clause for severable innovation. The result of the “but for" defense, focusing on a licensor’s incentive, can be supported by the background presented hereafter. In the first place, a major concern about a licensor’s strong market power may be very limited to the case where she is a monopolist or very close to being labeled as one. Rather, when a licensor is a relatively small firm, market competition is expected to work much better due to an equal footing enabled by grant-backs. More importantly, the second background is with regard to a licensee’s incentive to further innovate. If a licensee is a rational economic agent, she will weigh the advantage of using a grant-back clause (i.e. can employ an original patent to achieve innovation) against the disadvantage of not using it (i.e. cannot achieve innovation due to the absence of the patent), when contemplating the license contract that includes a grant-back clause. In other words, since a licensee incorporates the “pros and cons" concerning a grant-back clause into her decision-making process of concluding a license contract, the mere statement that grant-backs may potentially discourage a licensee’s incentive does not appear persuasive. As this paper indicates, it is likely that potential contracts can be concluded including a grant-back clause for severable innovation based upon the voluntary decision of a licensor and a licensee. Consequently, while a grant-back clause may be a measure of abuse of a licensor’s dominant position, it seems overemphasized that it discourages a licensee’s incentive to innovate for severable innovation.

There remains a further challenging investigation in this paper: the assumption that severable innovation entirely cancels out royalty payments from a licensee should be re-examined. Alternatively, we may need to assume that some royalty payments can accrue conditional upon the actual use of a base technology possessed by a licensor even for severable innovation. By this modified assumption, the robustness of the “but for" defense of a grant-back clause for severable innovation may be somewhat restricted.

2.7. Appendix: partially prohibited territorial restrictions

Let us reaffirm the robustness of the “but for " defense of a grant-back clause for severable innovation presented in Section 2.5 (two heterogeneous licensees) by investigating partially prohibited territorial restrictions. To this end, suppose that firm L is unable to establish a “virtual" territorial restriction with firm B due to, for example, antitrust policies prescribed by a competition authority. On the other hand, it is assumed that the territorial restriction between firms L and A is still maintained as long as concrete contracts aimed at innovation are concluded. In this regard, territorial restrictions are partially prohibited.

Under this presumption, since a contract offered by firm L in the absence of licensing a set of technologies is prohibited, Case 2-NN (NA; SN; SA) (firm L concludes a contract only with firm B) and Case 5-NA (SA) (firm L concludes contracts with firms A and B in Stage 1, but not with firm B in Stage 2) are non-existent. Note that the reservation profit of firm L in connection with the market of firm B decreases from $1-\frac{c}{2}$ to $1-c$. This operates to alter the profits of the firms, and hence, has a significant influence on the kind of contracts that are offered to firms A and B.

The detailed formulations and calculations have been omitted because they are almost identical to previous analyses presented. Table 2.12 exhibits the profits of the firms and social welfare. When a grant-back clause is available, the profit of firm L in Case 4-NA (SA) changes to $\pi_L^{NA4} = \pi_L^{SA4} = 3 + \frac{11\theta}{8} - \frac{5c}{4}$. Although the statement with regard to non-severable innovation is the same as before, the following proposition suggests that with regard to severable innovation, the previously described results should be slightly modified.

Non-severable innovation and non-availability of a grant-back clause (NN)

	Case 1-NN	Case 2-NN	Case 3-NN	Case 4-NN
π_L	$3-2c$	—	$3+\theta-\frac{3c}{2}$	$3+\frac{11\theta}{8}-\frac{5c}{4}$
π_A	0	—	$\theta+\frac{c}{2}$	$\frac{11\theta}{8}+\frac{3c}{4}$
π_B	0	—	0	$\frac{\theta}{4}+\frac{c}{2}$
W	$3-2c$	—	$3+2\theta-c$	$3+3\theta$

Non-severable innovation and availability of a grant-back clause (NA)

	Case 1-NA	Case 2-NA	Case 3-NA	Case 4-NA	Case 5-NA
π_L	$3-2c$	—	$3+\theta-\frac{3c}{2}$	$3+\frac{11\theta}{8}-\frac{5c}{4}$	—
π_A	0	—	$\theta+\frac{c}{2}$	$\frac{11\theta}{8}+\frac{3c}{4}$	—
π_B	0	—	0	$\frac{\theta}{4}+\frac{c}{2}$	—
W	$3-2c$	—	$3+2\theta-c$	$3+3\theta$	—

Note: This case is identical to “Severable innovation and availability of a grant-back clause (SA)".

Table 2.12. Profits and social welfare (partially prohibited territorial restriction).

Severable innovation and non-availability of a grant-back clause (SN)

	Case 1-SN	Case 2-SN	Case 3-SN		Case 4-SN	
			$\theta > c$	$\theta < c$	$\theta > c$	$\theta < c$
π_L	$3 - 2c$	—	$2 - \frac{c}{2}$	$2 + \frac{\theta}{2} - c$	$1 + \frac{c}{2}$	$1 + \frac{\theta}{2}$
π_A	0	—	$1 + 3\theta - \frac{3c}{2}$	$1 + \frac{3\theta}{2}$	$1 + 3\theta - c$	$1 + 2\theta$
π_B	0	—	0	0	$1 + \frac{c}{2}$	$1 + \frac{\theta}{2}$
W	$3 - 2c$	—	$3 + 3\theta - 2c$	$3 + 2\theta - c$	$3 + 3\theta$	$3 + 3\theta$

Table 2.12. Profits and social welfare (partially prohibited territorial restriction) (continued).

Proposition 2.6 (Severable innovation and two heterogeneous licensees in the presence of partially prohibited territorial restrictions between firms L and B) Severable innovation is assumed in what follows:

(1) Suppose that a grant-back clause is not available. Then, innovation is achieved and the improved technology is shared by firm L with firm A if $\frac{2}{3} < c < 1$ for $\theta > c$ and $\theta > 2(1 - c)$ for $\theta < c$. But the improved technology is never shared by firm B, and hence, first-best social welfare is not generated; and

(2) Suppose that a grant-back clause is available. Then, innovation is achieved and the improved technology is shared by all firms L, A, and B through the license contracts concluded between them, and hence, first-best social welfare is generated.

Proof (1) $\pi_L^{SN3} > \pi_L^{SN4}$ holds for every (c, θ) the same as before. We need to compare Case 1-SN with Case 3-SN. With regard to $\theta > c$, $\pi_L^{SN3} - \pi_L^{SN1} = (2 - \frac{c}{2}) - (3 - 2c) = -1 + \frac{3c}{2} > 0 \Leftrightarrow \pi_L^{SN3} > \pi_L^{SN1}$ if $\frac{2}{3} < c < 1$. In addition, with regard to $\theta < c$, $\pi_L^{SN3} - \pi_L^{SN1} = (2 + \frac{\theta}{2} - c) - (3 - 2c) = -1 + \frac{\theta}{2} + c > 0 \Leftrightarrow \pi_L^{SN3} > \pi_L^{SN1}$ if $\theta > 2(1 - c)$.

(2) Firm L prefers Case 4-SA rather than Case 3-SA since $\pi_L^{SA4} - \pi_L^{SA3} = (3 + \frac{11\theta}{8} - \frac{5c}{4}) -$

$(3+\theta-\frac{3c}{2}) = \frac{3\theta}{8}+\frac{c}{4} > 0 \Leftrightarrow \pi_L^{SA4} > \pi_L^{SA3}$. Let us check whether a grant-back clause is actually used by comparing the profits of firm L between Cases SA and SN. With regard to $\theta > c$, $\pi_L^{SA4}-\pi_L^{SN3} = (3+\frac{5\theta}{4}-\frac{5c}{4})-(2-\frac{c}{2}) = (1-\frac{3c}{4})+\frac{5\theta}{4} > 0 \Leftrightarrow \pi_L^{SA4} > \pi_L^{SN3}$. In addition, with regard to $\theta < c$, $\pi_L^{SA4}-\pi_L^{SN3} = (3+\frac{5\theta}{4}-\frac{5c}{4})-(2+\frac{\theta}{2}-c) = (1-\frac{c}{4})+\frac{3\theta}{4} > 0 \Leftrightarrow \pi_L^{SA4} > \pi_L^{SN3}$. Finally, we can confirm that irrespective of (θ, c), $\pi_L^{SA4}-\pi_L^{SN1} = (3+\frac{11\theta}{8}-\frac{5c}{4})-(3-2c) = \frac{5\theta}{4}+\frac{3c}{4} > 0 \Leftrightarrow \pi_L^{SA4} > \pi_L^{SN1}$. Hence, the profit of firm L in Case 4-SA is the highest of all cases. Since social welfare is $3+3\theta$ in Case 4-SA, first-best social welfare is achieved. ■

When territorial restrictions are partially prohibited between firms L and B, there may occur innovation even if a grant-back clause is not available for severable innovation. In particular, this result, being different from Lemma 2.6 which states that innovation is never achieved for severale innovation, stems from the fact that firm L can no longer reinforce the bargaining power relying on the territorial restriction with firm B. (The reservation profit decreases from $3-\frac{3c}{2}$ to $3-2c$.) Nevertheless, it is noteworthy that firm L has no incentives to transfer BT to firm B for severable innovation either, while the gain earned in the new market of firm B belongs to firm A. Again, grant-back clauses play a role in restoring the licensor's incentive to open her base technology. To conclude, as a grant-back clause succeeds in generating first-best social welfare, the “but for" defense is still robust.

References

[1] Bleeke, J. A. and Rahl, A. (1979), “The Value of Territorial and Field-of-Use Restrictions in the International Licensing of Unpatented Know-How: An Empirical Study", *Northwestern Journal of International Law and Business*, 1(2), pp. 450-491

[2] Choi, J. P. (2002), “A Dynamic Analysis of Licensing: The ‘Boomerang’ Effect and Grant-Back Clauses", *International Economic Review*, 43(3), pp. 803-829

[3] Cockburn, I. M. (2007), “Is the Market for Technology Working? Obstacles to Licensing Inventions, and Ways to Licensing Inventions, and Ways to Reduce Them", paper presented at the Conference on Economics of Technology Policy, Monte Verità, June 2007

[4] Delrahim, M. (2004), “The Long and Winding Road: Convergence in the Application of Antitrust to intellectual Property", U.S. Department of Justice, presented at the George Mason Law Review Symposium, October 2004

[5] European Union (2004), “Commission Note: Guidelines on the Application of Article 81 of the RC Treaty to Technology Transfer Agreements", *Official Journal of the European Union*, 2004/C 101/02, available at:
http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52004XC0427(01)

[6] European Union (2014), “Communication from the Commission: Guidelines on the Application of Article 101 of the Treaty on the Functioning of the European Union to Technology Transfer Agreements", *Official Journal of the European Union*, 2014/C 89/03, available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri= CELEX%3A52014XC0328(01)

[7] Japan Fair Trade Commission (2016), “Guidelines on the Antitrust Laws regarding the Use of Intellectual Property" (in Japanese), available at:
http://www.jftc.go.jp/dk/guideline/unyoukijun/chitekizaisan.html

[8] Moreira, S., Laursen, K. Leone, M. I., and Reichstein, T. (2012), “Cooperation or Competition: Grant-Back Clauses in Technology Licensing", paper presented at the DRUID Academy 2012, The Moeller Centre of University of Cambridge, January 2012, available at:
http://druid8.sit.aau.dk/acc_papers/ain1su9m8x2tjii0rtlonyktvidl.pdf

[9] OECD (2004), *Policy Roundtables: Intellectual Property Right*, Paper No. DAF/COMP(2004)24, Paris, available at
https://www.oecd.org/daf/competition/34306055.pdf

[10] Scotchmer, S. (1991), “Standing on the Shoulders of Giants: Cumulative Research and the Patent Law", *Journal of Economic Perspectives*, 5(1), pp. 29-41

[11] Scotchmer, S. (1996), “Protecting Early Innovators: Should Second-Generation Products Be Patentable?", *RAND Journal of Economics*, 27(2), pp. 322-331

[12] van Dijk, T. (2000), “License Contracts, Future Exchange Clauses, and Technological Competition," *European Economic Review*, 44(8), pp. 1431-1448

[13] Wäktare, E. (2007), “Territorial Restrictions and Profit Sharing Mechanisms in the Gas Sector: the Algerian Case", Competition Policy Newsletter, Issue 3, European Commission, pp. 19-21, available at:
http://ec.europa.eu/competition/publications/cpn/2007_3_19.pdf

[14] Zuniga, M. P. and Guellec, D. (2009), “Who Licenses out Patents and Why?: Lessons from a Business Survey", OECD Science, Statistical Analysis of Science, Technology and Industry Working Papers, Paper No. DSTI/DOC(2009)5, OECD, available at: http://www.oecd.org/science/inno/42477187.pdf

Chapter 3. Theory and Empirics of University Research and Teaching: Can We Simultaneously Increase University Research Output and Student Enrollment?

3.1. Introduction

Most universities not only conduct research to acquire universal knowledge but also teach students to enhance human capital, and these two activities exactly define what universities are. [1] Although a number of economics studies on universities have so far focused on the mechanisms of knowledge creation, the effects of knowledge diffusion from universities to industries, and the measures of promoting university research, [2] there have not been many studies investigating the complex interactions between their research and teaching activities. Meanwhile, we recognize the fact that research grant and tuition fee policies are highly likely to affect the achievement of both research output and student enrollment generated by universities.

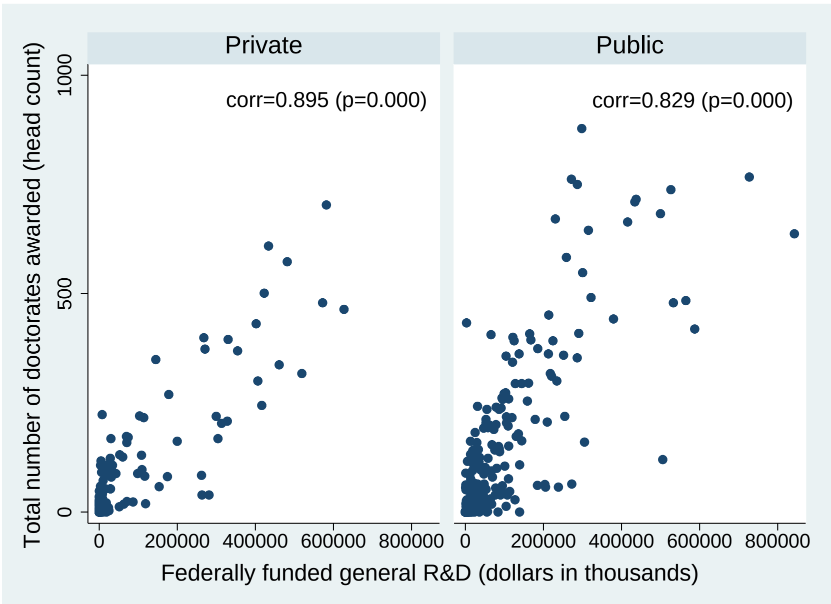
It is widely believed among developed countries that increased public research funds serve the direct purpose of producing more research output. [3] Indeed, on the

[1] Various definitions of universities have been presented. Haskins (1957) finds that modern universities have their roots in encouraging researchers to study disciplines in order to seek truth and knowledge. On the other hand, Mill (1867) indicates in his famed speech the importance of university education and the acquisition of specialist knowledge there.

[2] Representative empirical works of university knowledge diffusion include Jaffe (1989), Henderson et al. (1998), and Lach and Schankerman (2008). Also, Foray and Lissoni (2010) comprehensively survey the topics of university research and its knowledge creation.

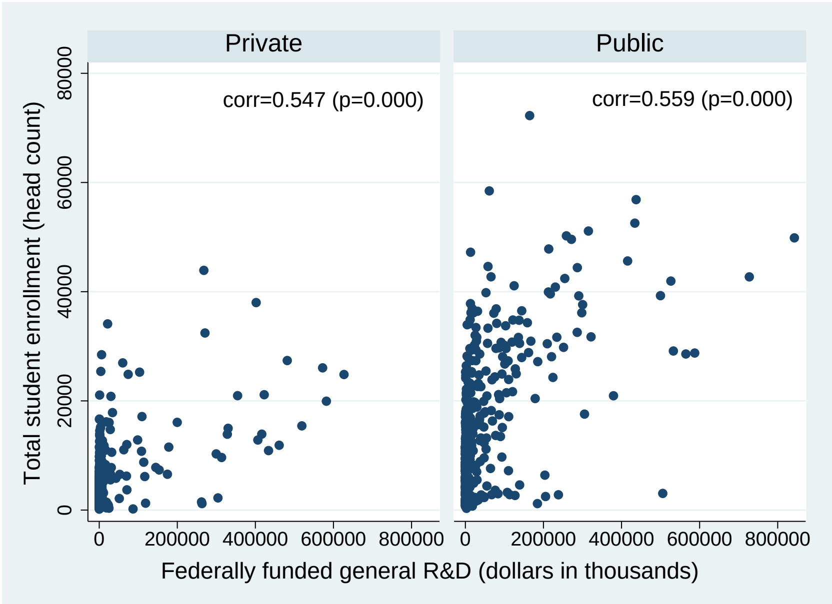
[3] According to OECD (2015), the level of research and development (R&D) conducted by OECD countries steadily rose by 2.7% in real terms from 2012 to 2013 along with the recovery from the global economic decline and financial crisis. This data seems to reflect policy makers' belief that R&D activities are essential for technological advancement and inclusive economic growth.

basis of cross-sectional data of U.S. universities and higher educational institutions (hereafter, “universities" as a whole) during the period of 2011, Figure 3.1 exhibits a strong positive correlation between the total number of doctorates awarded (which is used as a proxy of research output) and federally funded general R&D to both public and private universities. In addition, it appears from Figure 3.2 that total student enrollment is positively correlated with R&D expenditure of universities and that large universities with a great many students enrolled have a propensity to invest more in R&D than small universities.



Source: National Center for Science Engineering Statistics (NCSES).

Figure 3.1. U.S. federally funded general R&D versus number of doctorates awarded in 2011.



Source: National Center for Science Engineering Statistics (NCSES).

Figure 3.2. U.S. federally funded general R&D versus total student enrollment in 2011.

Certainly, we need to take note that these findings rely on a simple correlation analysis and only indicate average tendencies across universities, rather than any causal relationship. However, these facts provide us with motivation to further probe the detailed interactions between university research and teaching activities and their effects on research output and student enrollment. More precisely, the intriguing thing is to address how and in what circumstances research output and student enrollment are increased or decreased. In particular, the main objective of this chapter is to investigate theoretically how research output and student enrollment respond to changes in policy measures such as public research funds and tuition fee settings.

A few authors have addressed the issue of multi-tasking universities that face a tension between research and teaching activities. Del Rey (2001) analyzes a model where two competing universities conduct research and teaching in the same jurisdiction and they are financed by the government through a lump-sum amount and per-student allocation. She shows that depending on the parameters such as the finance scheme, teaching efficiency, and the relative weight of research activities, the model generates various combinations of equilibria; the universities conduct only teaching (or research) and teach selective (or mass) students. In the study by De Fraja and Iossa (2002), two universities are separately located and students choose whether and where to attend one university taking into account their mobility costs. Supposing that the prestige of each university relies on the number of students enrolled and its research outcome, they also derive several equilibrium configurations associated with varying mobility costs; for instance, in one equilibrium multiple universities accept students by adopting the same admission standards, whereas in another one university admits fewer high-achieving students by establishing a higher admission standard.

Although these studies regard research activities as “residual" in their total capacities, Beath et al. (2012) treat research as making a trade-off with teaching. By allowing universities to choose voluntarily the quality level of research and teaching, the authors demonstrate that when the government funding system is used as a tool to control research incentives of universities, a variety of university cultures may emerge, such as research-oriented and teaching-oriented universities. De Fraja and Valbonesi (2012), who construct a general equilibrium model of universities and students, compare research and teaching distributions among universities in accordance with university management policies: the unregulated private provision policy versus the government intervention policy. They argue that while the former policy inefficiently allows to spread research across all universities, the latter system can efficiently concentrate research and teaching on fewer more productive universities because the government intervention both improves research efficiency and assures teaching quality.

Lastly, from a different perspective, Gautier and Wauthy (2007) investigate an incentive problem within a university that needs to govern both research and teaching conducted by its individual departments and that redistributes aggregate tuition revenue to them. Positing that the departments are evaluated by the university based solely upon their research output, they find a trade-off problem. In other words, research activities can be increased due to yardstick competition caused by this assessment policy while teaching activities can be decreased due to free-riding by the departments that cannot appropriate their own tuition revenue. They also point out that when the departments are integrated into a multi-unit institution that has in nature a complementarity between research and teaching activities (i.e. strategic complementarity), both activities can be promoted.

Based on these existing studies, this chapter constructs a micro-founded university-student model with a trade-off between research and teaching activities of universities. The baseline model assumes that a single university intends to maximize its payoff (namely, academic prestige) obtained from research output given a fixed research fund that is allocated by a financing agency while conducting both research and teaching activities. Teaching is not the ultimate goal of the university, though it potentially contributes to its research budget through tuition revenue. Instead, by evaluating teaching offered by the university, innumerable students decide whether to attend the university if the benefit exceeds the cost. The university therefore needs to draw a fine balance between research and teaching activities in order to earn tuition revenue that also can be exploited as a research resource.

In particular, the capacity constraint, which imposes a limitation irrespective of monetary research resources on the total activities of the university, carries a critical meaning. This capacity may be regarded as a limitation of “ability" inherent to the university. Due to the existence of this constraint, the university is compelled to allocate appropriately its limited capacity to both research and teaching activities. Since it is usually difficult to enhance a university capacity in the short run, these two activities are jointly limited at some level. In this setting, if a university intends to increase its research activities, it may need to decrease its teaching activities instead (that is, a trade-off relation). In the quite long run, an improvement can happen that strengthens a university capacity, but we do not explicitly consider such a long-run effect in a theoretical framework and treat the capacity constraint as an exogenous parameter.

The findings of this paper are summarized in what follows. In the first place, the element of substitutability between research and teaching activities can be critical, in that one activity can increase the cost of another activity. If such substitutability is strong enough, not only student enrollment but also research output can be reduced as a result of an increase in a research fund. This seemingly paradoxical argument is deliberately demonstrated in a general model as a likely scenario.

Subsequently, assuming that the degree of substitutability is zero for analytical simplicity, this study points out that the results depend not only on whether the capacity of a university is fully utilized or not, but also on whether a tuition fee is fixed or controlled. More precisely, in the case of a *fixed* tuition fee, whereas research funds can increase both research output and student enrollment when the capacity is not fully used (i.e. “multiplier effect"), it crowds out student enrollment when the university operates at its full capacity (i.e. “crowding-out effect"). This former result is not so surprising, since research funds allow teaching activities, too, when the capacity constraint is slack. However, provided that the capacity constraint is binding as it is unnoticed, research funds reduce teaching activities because the university favors achieving more research output to get a higher payoff.

It is also revealed that when a tuition fee is *controlled* by a government authority in order to maximize tuition revenue, a small amount of a research fund is never expected to positively affect student enrollment due to the emergence of a “binary divide" among universities. This “binary divide" means multiple equilibria in accordance with capacity size; while a “large university" operates at its full capacity, a “small research institute" conducts marginal research and teaching activities. The mechanism is as follows. When the tuition fee is a controlled variable that is set for maximizing tuition revenue, it is optimal that the tuition fee rises in parallel with teaching activities because the decrease in student enrollment can be compensated by the enhancement of teaching. But since the positive effect on tuition revenue is relatively modest for the low-level teaching activities, the payoff of the university will decline as the teaching activities are augmented to some point and would then increase beyond that point. (In other words, mathematically, a saddle point emerges.) For this reason, if the capacity is small (large), it is rational for the university to select teaching activities at the minimal (maximal) level.

As a realistic extension, this study analyzes competition between two universities seeking students through their teaching activities in the same jurisdiction and compares the results with a single university, in each case of fixed and controlled tuition fees. It finds that under both schemes of tuition fees, such teaching competition has a possibility of increasing the aggregate of student enrollment, given that the total amount of research funds is evenly distributed to these universities. But it also indicates that research output may decrease under multiple universities since the advantage of increasing returns to scale in research output is nullified, which is benefited more by a single university.

Before conducting theoretical analyses, this chapter attempts an empirical investigation on the basis of U.S. university data. The empirical focus is primarily on examining dominant effects on research output and student enrollment in response to an increase in a research fund: substitutability versus complementarity between research and teaching activities; or multiplier versus crowding-out effects. This analysis divides the total number of universities (including other higher educational institutions) under the headings of public and private universities, which seem to have different natures from the other university categories. In spite of some limitations of the analysis, it is suggested that research funds may be negatively related not only to student enrollment but also to research output of private universities. This result implies that substitutability between research and teaching activities (and the crowding-out effect) may be the dominant force for private universities. The following discussions on theoretical parts are motivated by this remarkable empirical finding.

While sharing some similarities with earlier works, this study differs from them on some points. First, although the model includes only a single university in the baseline model in contrast to other studies (Del Rey, 2001; De Fraja and Iossa, 2002; De Fraja and Valbonesi 2012), we obtain some new findings regarding the effect of a research fund on both research output and student enrollment. Second, this study allows students to endogenously make their own decisions in a multi-stage game as to whether they attend a university considering the level of teaching, so that their decisions affect the research and teaching activities of the university. Although most existing studies incorporate students as necessary agents into their models, some deal with students as passively responding to constraints set by universities. Third, this study also focuses on a tuition fee per student that is a key source of university revenue, while other studies have not conducted such a thorough investigation. It is noticeable to distinguish the case where a tuition fee is exogenously fixed from the case where it is endogenously controlled in order to maximize tuition revenue. The analyses of these two cases illustrate that the implications of providing a research fund can be entirely different between them. Finally, this study explicitly considers a capacity limitation of the university to undertake research and teaching activities. It is made clear that this capacity limitation influences the action taken by a university in conjunction with the above tuition fee schemes.

In the meanwhile, this chapter omits some important aspects which other authors draw attention to in their models. More precisely, Beath et al. (2012) and Gautier and Wauthy (2007) relate the distribution of research funds to research productivities (i.e. abilities) of universities. The productivity issue is not considered because the model posits a single university or multiple homogeneous universities. It is also assumed that there is no competition among universities seeking research funds provided by a financing agency; instead, they compete for students based on their teaching activities. Admittedly, this assumption reflects only a part of the realities surrounding universities. However, since it is obvious that research funds should be preferentially allocated to the most research productive universities under many general scenarios, the allocation problem of research funds has not been addressed in this analysis. [4] We focus exclusively on the effect of competition for student applicants, which allows us to conduct a more simple analysis of the interplays between university research and teaching activities under the pressure of competition.

Moreover, this chapter sets aside distinct issues that are related to organizational economics of universities in order to narrow the research focus to interactions between research and teaching activities within a university. It highlights a university itself as an independent institution. The internal organization or incentives of universities (Beath et al., 2003; Caballero et al., 2004; Gautier and Wauthy, 2007) and their motivations compared to private firms (Aghion et al., 2008; Lacetera, 2009) are not

[4] A number of governments have recently introduced “competitive research funding" mechanisms where universities obtain future research resources based on the evaluation of their research output. It seems that by establishing a competitive environment for research activities, governments intend to elicit information about research productivities from universities and to encourage productivity improvements on a long-term basis.

investigated. Although the process of idea generation (Hellman and Pertotti, 2011; Scotchmer, 2013) and the relationship between basic and applied research (Beath et al., 2003; Banal-Estañol and Macho-Stadler, 2010; De Fraja, 2012) have been investigated, we do not address these issues either. These topics are highly interesting to examine, but they are eliminated as out of scope of the study.

The following briefly introduces the reminder of this chapter. At the first outset, Section 3.2 describes an empirical implication using U.S. university data. Section 3.3 outlines a model structure and describes the decisions of a university and students. Section 3.4 derives theoretical results in the case of a single university. Section 3.5 assumes a controlled tuition fee and compares the results with those derived from a fixed tuition fee. Sections 3.4 and 3.5 first introduce a general model framework, and subsequently present illustrative cases. Section 3.6 investigates an extended model where two universities compete for student enrollment in the same location. Finally, Section 3.7 makes concluding remarks including policy implications. All the mathematical proofs and the detail of the empirical analysis are compiled in Section 3.8.

3.2. Empirical implications

3.2.1. Overview of the analysis

Section 3.2 aims to find empirical implications by using an econometric method so as to guide some theoretical results that are derived in the later sections. The objective of this empirical analysis is to find what effects an increased research fund has on research output and student enrollment based on regression analyses that include other parameters such as a tuition fee. Although this chapter exhibits many theoretical results regarding the scheme of tuition fee settings and teaching competition between multiple universities, the simple empirical analysis presented here does not dwell on these differences, which is not the ultimate goal of this section. Hence, the following investigations are mere a preliminary “empirical challenge" and does not intend to thoroughly probe every theoretical result discussed afterwards.

Figures 3.1 and 3.2 presented that federally funded general R&D is positively correlated with the number of doctorates awarded (the proxy of research output) and total student enrollment, respectively, for U.S. public and private universities. In view of these relationships, the theoretical analyses could include the following possibilities: a positive or negative effect. (These effects are brought about in accordance with the conditions such as the degree of substitutability between research and teaching activities and the capacity constraint of a university. The later theoretical sections will give a detailed descriptions to them.)

First, since a research fund normally plays a role in increasing research output, a positive effect is anticipated (cf. Payne and Siow, 2003; Beaudry and Allaoui, 2012; Yonetani et al. 2013). As the theory will suggest, if the degree of substitutability between research and teaching activities is zero, or if it is small enough, research output is likely to increase in response to an increased research fund. On the other hand, the theory also will suggest that if substitutability is sufficiently strong, research output can deteriorate by an increase in the research fund. The last prediction seems quite paradoxical, but is of interest to be examined from the standpoint of empirical data, too. Second, just as importantly, we will see that the effect of an increased research fund on student enrollment can be both positive and negative in accordance with various situations. Whereas the positive effect is due to weak substitutability (or complementarity) between research and teaching activities, or the multiplier effect induced under non-binding capacities, the negative effect is due to strong substitutability between the two activities, or the crowding-out effect caused by full capacities. (Multiplier and crowding-out effects will be explained later.) Unfortunately, if we obtain significant signs from the regression analysis, it is not easy to identify each factor, i.e. complementarity or multiplier effect; substitutability or crowding-out effect. Therefore, all we can do is to state a conjecture on that cause in the underlying background.

Tuition fee (undergraduate and graduate) and the number of academic staff are also included as control variables. A rise in a tuition fee theoretically leads in two conflicting directions: a positive effect under inelastic student enrollment against a tuition fee (that is, the uncompetitive student market exemplified by a single university) and a negative effect caused by competition between universities that seek students. Although this examination itself is intriguing compared to existing research, [5] it is not our main focus. The number of academic staff is regarded as a proxy of a university

[5] Helmet and Marcotte (2011) estimate that an increase in 100 U.S. dollars in a tuition fee decreases student enrollment by about 0.25% based on all U.S. public universities in the period of 1991 to 2006. Using the system estimation that considers both demand and supply sides for university education, Neill (2009) reveals that enrollment demand in Canadian provinces declines by 2.5-5% with a 1000 Canadian dollars increase in a tuition fee even with an improvement in student financial aid. Furthermore, Hübner (2012), who employs introduction of tuition fee in the German states in 2007 as a natural experiment, also finds a negative effect of a tuition fee rise on student enrollment probability. By contrast, Canton and de Jong (2005) suggest that students are not responsive to tuition fee, but financial support, education premium on future labor market earnings, and alternative wage as an outside option, by investigating the post-war dataset in the Netherlands.

capacity, while it may represent the activity level of research and teaching. The positive sign of this variable is normally expected if capacity constraints of universities are critical.

The empirical data is based on U.S. academic institutions and science database (*National Center for Science Engineering Statistics* [*NCSES*]). The regression analyses are conducted for the number of doctorates awarded (total and natural science) and total student enrollment, respectively. With regard to independent variables, federally funded general R&D (hereafter, simply “general R&D") or federally and locally (state-governmentally) funded science and engineering R&D (hereafter, simply “science R&D") are mainly utilized. Although general and science R&D are not necessarily the exact measures of public research funds, they are considered adequate proxies. Finally, regression analyses are conducted in accordance with the samples being divided into threes categories: public universities; private universities; and both public and private universities combined as a whole.

A more detailed argument about the dataset, estimation methods, and the results are delegated to Subsection 3.8.5 in Appendices.

3.2.2. Summary and discussions

The bottom-line results of empirical analyses are as what follows:

1. Whereas publicly funded general and science R&D is positively related to the total and science number of doctorates awarded by public universities, it is not or might be negative for private universities. This result suggests that research

and teaching activities may be strong substitutes for private universities.

2. Publicly funded science R&D may have negative relations with total student enrollment for private universities, although the result is not robust. This implies that research and teaching activities may be substitutes or that total student enrollment may be crowded out by public research funds for private universities.

3. The undergraduate tuition fee may be negatively associated with total student enrollment for public universities.

4. The number of academic staff may be positively correlated with the number of science doctorates awarded by private universities. It may also have positive relations with total student enrollment for both public and private universities.

With regard to Result 1, it is striking that publicly funded R&D has no effects or may have negative effects on the number of doctorates awarded by private universities (more precisely, general R&D may have negative effects on the number of science doctorates awarded) possibly due to strong substitutability between research and teaching activities. Although there is a discussion about whether the number of doctorates awarded is a reliable indicator of research output, the result that research and teaching activities may be substitutes for one another would still be the case. On the basis of this finding, policymakers and university administrative officials need to recognize that research funds could be for some situations ineffectual in increasing the number of doctorates awarded, and hence, to design effective university systems that can activate research output of universities.

Let us focus on Result 2. There are two theoretical reasons for the negative effects of publicly funded R&D in case of student enrollment: substitutability between research and teaching activities, and the crowding-out effect. Although it is difficult to disentangle between the two causes, both are likely in the case of private universities. Private universities (in particular, small educational colleges) often place an emphasis on teaching activities in their action policies, and in such a situation, the degree of substitutability between research and teaching activities is further strengthened. In addition, since private universities tend to be small, their limited capacities can be serious obstacles to increasing their teaching activities. Result 4 seems to provide some support of this view by proving that the estimates of the number of academic staff are significantly positive and the magnitudes are relatively large in the case of private universities. It is hence suggested that capacities or scales are an important factor for student enrollment to be generated particularly at small universities.

And yet, I am afraid to say that we have enough difficulty identifying which impact, substitutability or the crowding-out effect, is actually prevalent. Furthermore, it is also difficult to have a clear understanding of what types of the crowding-out effect work in the theoretical analyses. (The theoretical implications could be different for fixed or tuition fee, respectively.) To sum up, if we take the position that university management policies need to address student enrollment as well as research output, it is important to direct more of our attention toward these possible negative effects caused by public research funds.

Finally, Result 3 indicates that competition for total student enrollment may be in force for public universities. It follows from a different viewpoint that private universities may be possibly more differentiated than public universities in undergraduate teaching, such as education styles and a wide selection of various teaching courses offered, in particular, those offered by educational colleges. But care must be taken in the interpretation of this empirical finding as a lack of robustness is evident.

Inspired by these empirical implications, we hereafter develop some theoretical models in order to probe the question, “can we simultaneously increase university research output and student enrollment?"

3.3. Basic model outline

The objective of this section is to formulate and analyze a general university-student model before deriving the theoretical results based on a specific parameterized model, which is described in Section 3.4. This general model is expected to provide a favorable outlook of the theoretical results throughout this chapter.

3.3.1. Players

In a particular jurisdiction, there exists a single university and numerous prospective students. (Section 3.6 investigates the case of multiple universities.) The details of how each player behaves in this model is described below.

University

In order to conduct its activities, a university needs to input positive research and teaching effort, $r > 0$ and $t > 0$, respectively, represented as $e = (r, t)$. Research and teaching effort can be interpreted as activity levels of the university, for example, improving research environments and training teaching staff, respectively.

The university is assumed to be constrained by a finite capacity, $\bar{a} > 0$, concerning effort, the level of which is defined by $r + t \leq \bar{a}$ for any r and t. [6] (The “bar" denotes a capacity limitation of the university.) In other words, the capacity exogenously specifies an upper bound of total effort, and thus, it is also regarded as an inherent “ability" of the university. This assumption of a capacity limitation can be justified in the short run rather than in the long run for the following reason. That is, irrespective of whether monetary resources are abundant, it is difficult to immediately enhance a university capacity in the short run, for example, by constructing new campuses or hiring more highly-qualified university faculty members, which are all likely to reinforce the intrinsic ability of the university. Hereafter, we particularly focus on the short-run framework where the capacity limitation is fixed. In the theoretical part, though not analyzing implications of a long-run flexible change in the capacity, the model intends to derive conditions of a capacity scale that is required to produce desired outcomes in research and teaching activities when a tuition fee is controlled. Rightfully, this implies that the improvement in a capacity matters in the long-run framework.

[6] As defined in the cost function later, it is also possible to state that the cost becomes infinite beyond the capacity level, $\bar{a}$.

Research output, R, is determined by both research effort, and the total budget that the university can make readily available for research: $R = R(r, b)$, where b is a research budget. Although Gautier and Wauthy (2007) considers research efficiency in addition to research effort that can both affect research output of the university, our analyses abbreviate the former element for the purpose of simplicity. [7] Suppose that research effort is separable from the research budget and that the research output function is determined by the simple product of the two: $R = rb$. Although the research output function can be assumed to be diminishing returns to scale for the research budget, [8] this allows us to greatly simplify the following analyses without losing the essence of the discussion. (This specification also enables us to obtain a closed-form solution in the illustrative case.) Or the capacity limitation, $\overline{a}$, can be viewed as supplementing the assumption of diminishing returns to scale.

Additionally, the budget, b, consists not only of a research fund but also of tuition revenue. It is hence denoted by $b = F + sn$, where $F > 0$ is a research fund allocation, s is a tuition fee per student, and n is the number of students enrolled. Accordingly, other things being equal, higher student enrollment could raise research output through an increase in the research budget of the university.

Research and teaching activities inevitably involve costs, such as establishing experimental instruments in research labs and hiring professional teaching staff. The

[7] In reference to Gautier and Wauthy (2007), the research productivity of a university can be represented as $\theta = \theta(r, \rho)$, where ρ is research efficiency. It is natural to think that $\frac{\partial\theta}{\partial r} > 0$ and $\frac{\partial\theta}{\partial\rho} > 0$. In this model, the research productivity is of the exact equivalence to the amount of research effort.

[8] Gautier and Wauthy (2007) posits a strictly concave function regarding a university budget such that $V(b) = b^{1-\alpha}$ with $0 < \alpha < 1$ in order to probe a difference in allocated research funds across departments. Note that concavity of the payoff function of the university in our model is still guaranteed because the cost function is assumed to be convex as described later.

cost function is described by $C = C(r, t)$ with $\frac{\partial C}{\partial r} > 0$ and $\frac{\partial C}{\partial t} > 0$, [9] which is typically assumed continuous and higher-order differentiable at any points. It is postulated that $C(r, t)$ is a strictly convex function: $\frac{\partial^2 C}{\partial r^2} > 0$, $\frac{\partial^2 C}{\partial t^2} > 0$, and $(\frac{\partial^2 C}{\partial r^2})(\frac{\partial^2 C}{\partial t^2}) - (\frac{\partial^2 C}{\partial r \partial t})^2 > 0$. This condition is that the Hessian matrix of $C(r, t)$ is a positive definite. It is also assumed that $\frac{\partial^3 C}{\partial r^3} = \frac{\partial^3 C}{\partial t^3} = 0$, which implies that the order of $C(r, t)$ regarding r and t is a maximum of two. The sign of the cross derivative with regard to research and teaching effort, $\frac{\partial^2 C}{\partial r \partial t}$, is not obvious, depending on whether the effort is a substitute, complement, or independent. If the effort is a substitute (complement) in terms of the cost function, we can maintain that $\frac{\partial^2 C}{\partial r \partial t} > 0 (< 0)$. That is, research and teaching activities being reciprocally substitutes (complements) suggests that an increase in one activity raises (or lowers) the marginal cost of the other, so that there exists negative (positive) externalities between them.

The payoff (utility) of the university is assumed to be determined by the value of its research output minus its cost of effort:

$$U(r, t) = wR(r, b) - C(r, t) = wrb - C(r, t), \quad \text{where } b = F + sn. \tag{3.1}$$

w implies the physical or mental “reward" obtained from one unit of research output (e.g. patent license fee, emotional satisfaction, etc.). In the later analysis, the price of research output is normalized as $w = 1$ for analytical simplicity. It is also reasonable to postulate that the revenue must exceed the cost of research and teaching effort, and

[9] One may suspect that the cost function depends on the number of students, n, too. In fact, student enrollment seems to positively affect the cost by way of a “congestion effect"; for example, the more students attend, the more the university is required to reinforce teaching staff and school buildings. However, as explained later, the number of students is determined by teaching effort, and as a result, the cost function is also considered being indirectly linked to student enrollment.

therefore, $U(r,t) \geq 0$ must be guaranteed.

As De Fraja and Valbonesi (2012) point out, the ultimate goal of universities is assumed to achieve so-called academic “prestige" by producing research output, but not teaching outcome. Our modeling assumption is also based upon their idea. [10] And yet, this assumption does not necessarily means that universities underrate teaching activities. It is rather contemplated in this analysis that universities may view teaching activities as indirectly affecting its prestige by increasing their research budgets to be utilized for research activities. On the basis of this simple model, the university maximizes the payoff function given by Equation (3.1) with regard to both research and teaching effort (*r* and *t*, respectively).

Students

The decision to enroll made by students rests on teaching effort of a university. [11] The reason for this is that if students accumulate a sufficient amount of human capital through quality university teaching, they can gain an advantage in hope of obtaining better jobs after graduating from the university. Better teaching is likely to promote students’ fundamental abilities and professional skills that can be applied to future business activities. Or students can signal to employers information about their com-

[10] The U.S. evaluation system of higher education provides universities with strong incentives to undertake outstanding research, because they are heavily dependent on external funds in a greater part of their research budgets and competing for research grants provided by central and regional governments and private companies. In this system, if a university does not produce satisfactory research output, it cannot gain any research funds and faces a difficulty of continuing its research activities.

[11] For analytical reasons, like research efficiency, this study also omits the aspect of “teaching efficiency" that can be closely connected with the quality of teaching.

petencies obtained through university education compared to other students. [12]

A mobility cost, $k > 0$, and a tuition fee, $s > 0$, are also highly likely to affect the students' decisions, so that student enrollment, $n = n(t, k, s)$, is assumed. Notice that we do not have to literally interpret k as indicating some physical distance between the location of a particular student and the university. Rather, it could be that k represents the difficulties of entrance examination or of getting caught up with their studies after gaining admittance, both of which impose some kind of psychological burden on students. In this regard, the students are horizontally differentiated by a mobility cost as to whether they intend to attend the university or not. A tuition fee, s, is normally expected to have a negative impact on student enrollment. If the tuition fee of the university rises, students will cease admission to the university or chooses to attend another university outside the jurisdiction. Finally, assuming that $n(t, k, s)$ is continuous and higher-order differentiable at any points, we postulate $\frac{\partial n}{\partial t} \geq 0$, $\frac{\partial n}{\partial k} \leq 0$, $\frac{\partial n}{\partial s} \leq 0$, $\frac{\partial^2 n}{\partial t^2} \leq 0$, $\frac{\partial^2 n}{\partial t \partial k} \leq 0$, and $\frac{\partial^2 n}{\partial t \partial s} \leq 0$.

Financing agency

It is assumed that a governmental financing agency allocates a constant amount of a research fund, F, to the university. In a single-shot research model as considered here, there are no links in this model between consequential research output and future research funds provided by the financing agency. This does not suggest that a university increases research output simply expecting richer research funds in subsequent peri-

[12] For instance, it is the standard for large employers in the UK to require a "2:1" (first or upper second class honors) when offering students jobs or even interviews. In reference to the effect of school quality and educational performance on earnings of graduates, see Rumberger and Thomas (1993).

ods to follow based on the evaluation of previous research achievements. We have to also take note that since only a single university exists at the current point, there is no allocation problem with research funds.

3.3.2. Timing of the model

The model framework is described on the basis of a multi-stage game. The timing of the model is in what follows.

0. The variables (w, k, s) are realized exogenously. [13] The financing agency allocates a constant amount of a research fund, F, to the university.

1. The university chooses the level of research and teaching effort, $\boldsymbol{e} = (r, t)$, respectively.

2. Students choose whether or not to attend the university.

3. The payoffs of the university and students are realized.

The game is solved by backward induction to find a subgame-perfect equilibrium.

3.3.3. Equilibrium solution

Based on backward induction, the university maximizes its payoff given that student enrollment is obtained in Stage 2:

$$\max_{r,t} U(r, t) = rb - C(r, t) \tag{3.2}$$

$$\text{subject to } b = F + sn(t, k, s). \tag{3.3}$$

[13] Section 3.5 deals with the tuition fee, s, as a control (endogenous) variable.

From the above Equations (3.2) and (3.3), the payoff function of the university is abbreviated into $U = r[F + sn(t, k, s)] - C(r, t)$. On the assumption of a positive interior solution (i.e. $\boldsymbol{e}^* = (r^*, t^*) > \boldsymbol{0}$ and $r^* + t^* < \overline{a}$ are satisfied), the first-order condition is formulated as follows:

$$\frac{\partial U(r,t)}{\partial r} = F + sn(t,k,s) - \frac{\partial C(r,t)}{\partial r} = 0, \tag{3.4}$$

$$\frac{\partial U(r,t)}{\partial t} = rs\left[\frac{\partial n(t,k,s)}{\partial t}\right] - \frac{\partial C(r,t)}{\partial t} = 0. \tag{3.5}$$

$\boldsymbol{e}^* = (r^*, t^*)$ satisfies both Equations (3.4) and (3.5). From hereon, the variables of the functions are abbreviated for descriptive simplicity. To secure a global maximum solution, we intend to confirm whether the second-order condition is satisfied using a Hessian matrix of $U(r, t)$:

$$\tilde{U} = \begin{bmatrix} \frac{\partial^2 U}{\partial r^2} & \frac{\partial^2 U}{\partial r \partial t} \\ \frac{\partial^2 U}{\partial t \partial r} & \frac{\partial^2 U}{\partial t^2} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 C}{\partial r^2} & s(\frac{\partial n}{\partial t}) - \frac{\partial^2 C}{\partial r \partial t} \\ s(\frac{\partial n}{\partial t}) - \frac{\partial^2 C}{\partial r \partial t} & -\frac{\partial^2 C}{\partial t^2} \end{bmatrix}. \tag{3.6}$$

From the assumption of the cost function, we obtain $-\frac{\partial^2 C}{\partial r^2} < 0$, $-\frac{\partial^2 C}{\partial r^2} < 0$. If it is posited that the determinant of $\tilde{U}$ is positive like $|\tilde{U}| = (\frac{\partial^2 C}{\partial r^2})(\frac{\partial^2 C}{\partial t^2}) - [\frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t})]^2 > 0$, [14] the payoff function is strictly concave, and hence, $\boldsymbol{e}^* = (r^*, t^*)$ induces a global maximum. Moreover, equilibrium student enrollment and research output are defined as $n^* = n(t^*, k, s)$ and $R^* = r^*(F + sn^*)$, respectively. In order to observe a change in the endogenous variables in an interior point, we suppose that under-enrollment of students occurs. (That is, some students do not apply for admission to the university.)

[14] When positive substitutability concerning the cost function, $\frac{\partial^2 C}{\partial r \partial t} > 0$, is assumed, $|\tilde{U}| > 0$ is always satisfied from the assumption made with $C(r, t)$ and $n(t, k, s)$. This is demonstrated by $|\tilde{U}| = (\frac{\partial^2 C}{\partial r^2})(\frac{\partial^2 C}{\partial t^2}) - [\frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t})]^2 > (\frac{\partial^2 C}{\partial r^2})(\frac{\partial^2 C}{\partial t^2}) - (\frac{\partial^2 C}{\partial r \partial t})^2 > 0$.

3.3.4. Comparative statics of a research fund

An interesting undertaking is to analyze the effect of a research fund on research output and student enrollment when substitutability (or complementarity) exists. In other words, as the subtitle of this paper poses, “Can we simultaneously increase research output and student enrollment?". From this standpoint, let us direct our attention to investigating comparative statics of research output and student enrollment with regard to an increase in a research fund. (See Subsection 3.8.2 for the comparative statics with regard to a mobility cost and a tuition fee.) To this end, we take the derivatives on both sides of Equations (3.4) and (3.5) by F, respectively:

$$\begin{aligned}
&1 + s\left(\frac{\partial n}{\partial t}\right)\frac{\partial t^*}{\partial F} - \left[\left(\frac{\partial^2 C}{\partial r^2}\right)\frac{\partial r^*}{\partial F} + \left(\frac{\partial^2 C}{\partial r \partial t}\right)\frac{\partial t^*}{\partial F}\right] = 0 \\
&\Longleftrightarrow \left(\frac{\partial^2 C}{\partial r^2}\right)\frac{\partial r^*}{\partial F} + \left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\frac{\partial t^*}{\partial F} = 1, \qquad (3.7) \\
&s\left(\frac{\partial n}{\partial t}\right)\frac{\partial r^*}{\partial F} + rs\left(\frac{\partial^2 n}{\partial t^2}\right)\frac{\partial t^*}{\partial F} - \left[\left(\frac{\partial^2 C}{\partial r \partial t}\right)\frac{\partial r^*}{\partial F} + \left(\frac{\partial^2 C}{\partial t^2}\right)\frac{\partial t^*}{\partial F}\right] = 0 \\
&\Longleftrightarrow \left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\frac{\partial r^*}{\partial F} + \left[\left(\frac{\partial^2 C}{\partial t^2}\right) - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right]\frac{\partial t^*}{\partial F} = 0. \qquad (3.8)
\end{aligned}$$

The matrix notation of Equations (3.7) and (3.8) is provided by:

$$\begin{bmatrix} \frac{\partial^2 C}{\partial r^2} & \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) \\ \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) & \frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2}) \end{bmatrix} \begin{bmatrix} \frac{\partial r^*}{\partial F} \\ \frac{\partial t^*}{\partial F} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \qquad (3.9)$$

Let us denote the first matrix of Equation (3.9) as A_F. Its determinant is calculated as: $|A_F| = (\frac{\partial^2 C}{\partial r^2})[\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})] - [\frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t})]^2$. In this analysis, while $|A_F| > 0$ is assumed, it is always satisfied when $\frac{\partial^2 C}{\partial r \partial t} > 0$ holds (see also Footnote 13 for this argument). By the use of the “Cramer’s rule", we can obtain the following solution of the simultaneous equations:

$$\frac{\partial r^*}{\partial F} = \frac{\begin{vmatrix} 1 & \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) \\ 0 & \frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2}) \end{vmatrix}}{|A_F|} = \frac{1}{|A_F|}\left[\frac{\partial^2 C}{\partial t^2} - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right] > 0, \tag{3.10}$$

$$\frac{\partial t^*}{\partial F} = \frac{\begin{vmatrix} \frac{\partial^2 C}{\partial r^2} & 1 \\ \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) & 0 \end{vmatrix}}{|A_F|} = \frac{1}{|A_F|}\left[s\left(\frac{\partial n}{\partial t}\right) - \frac{\partial^2 C}{\partial r \partial t}\right]. \tag{3.11}$$

Clearly from Equation (3.10), an increased research fund always produces more research effort. In addition, if research and teaching activities are complements (i.e. $\frac{\partial^2 C}{\partial r \partial t} < 0$), we necessarily obtain $\frac{\partial t^*}{\partial F} > 0$ from Equation (3.11). We can also see that when there is no substitutability nor complementarity between research and teaching activities (i.e. $\frac{\partial^2 C}{\partial r \partial t} = 0$), the result that an increased research fund enhances teaching effort, $\frac{\partial t^*}{\partial F} > 0$, still holds. More interestingly, if the degree of substitutability is large enough such that $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t})$, we derive $\frac{\partial t^*}{\partial F} < 0$, which implies that teaching effort declines in response to the increase in a research fund. The important point here is that this negative effect on teaching effort could reduce not only student enrollment but also research output in an extreme case. Therefore, the answer of the question at the beginning of this subsection is usually the case, but in a particular situation, it could be difficult. Proposition 3.1 summarizes the results of comparative statics regarding a change in a research fund.

Proposition 3.1 With regard to the effect of an increased research fund, F, on research and teaching activities, we can obtain:

(1) $\frac{\partial r^*}{\partial F} > 0$ holds for any $\frac{\partial^2 C}{\partial r \partial t}$ (irrespective of substitutability or complementarity);

(2-i) $\frac{\partial t^*}{\partial F} > 0$, $\frac{\partial n^*}{\partial F} > 0$, and $\frac{\partial R^*}{\partial F} > 0$ for $\frac{\partial^2 C}{\partial r \partial t} < s(\frac{\partial n}{\partial t})$;

(2-ii) $\frac{\partial t^*}{\partial F} < 0$ and $\frac{\partial n^*}{\partial F} < 0$ for $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t})$; and

(3) $\frac{\partial R^*}{\partial F} < 0$ for $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t}) + \Omega$, where $\Omega = \frac{|A_F| r^* + [\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})](F + sn^*)}{r^* s(\frac{\partial n}{\partial t})} > 0$.

It is natural that an increased research fund entices the university to induce greater research effort due to the enriched total research budget (Proposition 3.1 [1]). Because of this additional research budget, the university normally finds it more profitable to devote more teaching effort, too, if the degree of substitutability is small enough (Proposition 3.1 [2-i]). It is also expected that as student enrollment increases, the research budget gets larger due to increased tuition revenue, and as a result, the university is likely to produce more research output.

But the effects of a research fund on teaching effort and student enrollment are not uniform according to substitutability between research and teaching activities. If the substitutability is strong enough that it brings about additional unwanted costs, teaching effort is decreased by contraries and then followed by a decline in student enrollment (Proposition 3.1 [2-ii]). Furthermore, the result of Proposition 3.1 (3) is paradoxical and much more controversial; an increase in a research fund could lead to a decrease in research output, which entirely contradicts the common notion envisioned by many individuals including policymakers. The intuition is explained as follows. When research and teaching activities are reciprocally strong substitutes as mentioned before, the decrease in teaching effort leads to reduced student enrollment. Since tuition revenue earned from students is also greatly reduced, there could be less research output produced as a result due to the even smaller research budget that is appropriated for research activities. Consequently, it is theoretically demonstrated that for strong substitutability, an increased research fund could decrease research output produced by the university. This Proposition 3.1 (3) is a seemingly paradoxical result, and yet, the theory suggests that it could actually happen and is also very indicative.

Repeatedly, Section 3.2 examined from an empirical perspective Proposition 3.1, as to whether such a negative relationship between a research fund and research output and student enrollment is observed or not. As we have seen, although we need to carefully interpret the empirical findings, a research fund could negatively affect not only student enrollment but also research output for U.S. private universities.

3.4. Modeling of an illustrative case

3.4.1. Analysis when substitutability exists

In Section 3.4, we investigate an illustrative case to derive explicit solutions for the model by parameterizing the formulations that have been specified above. The following parameterization is a mere benchmark, and not entirely general; however, since it satisfies some important qualitative nature of the model, the parameterization certainly allows us to illustrate the general behavior of the model.

Let us move back to the decision made by a university. As the previous section, the research output function is defined as $R = rb$ where $b = F + sn$. We suppose that the cost function takes a form of $C(r, t) = \frac{r^2}{2} + \frac{t^2}{2} + \varepsilon rt$. This choice encompasses the factors that lead to convexity regarding research and teaching effort (i.e. $\frac{r^2}{2}$ and $\frac{t^2}{2}$), but at the same time, allows research and teaching effort to affect the cost through the interaction term, εrt. The element, ε, represents the degree of substitutability (or complementarity) between university research and teaching effort in terms of the cost function. More precisely, while research and teaching effort is mutually a substitute for $\varepsilon > 0$, it is mutually a complement for $\varepsilon < 0$. This cost function satisfies the previous conditions assumed in Section 3.3: $\frac{\partial C}{\partial r} = r > 0$, $\frac{\partial C}{\partial t} = t > 0$, $\frac{\partial^2 C}{\partial r^2} = \frac{\partial^2 C}{\partial t^2} = 1 > 0$, and $\frac{\partial^3 C}{\partial r^3} = \frac{\partial^3 C}{\partial t^3} = 0$. But $(\frac{\partial^2 C}{\partial r^2})(\frac{\partial^2 C}{\partial t^2}) - (\frac{\partial^2 C}{\partial r \partial t})^2 = 1 - \varepsilon^2 > 0$ for $-1 < \varepsilon < 1$ is also required for the convexity of $C(r, t)$ (that is, ε needs to be bounded).

Next, in a student market, each student decides whether to apply for admission given teaching effort, t, made by the university. The choice by a student is modeled as follows. Students are assumed to be evenly distributed over the horizontal line, the length of which is normalized to 1. Meanwhile, the university is located at the middle point ($\frac{1}{2}$) of this line. In Hotelling's model (Hotelling, 1929), players such as firms or shops determine their own locations in order to differentiate their products from others. By way of contrast, our model postulates that the location of the university is fixed at the middle point, which implies that the university is assumed to be situated in a balanced place within the jurisdiction. (The university exists in the center of the jurisdiction.) In addition, since it is not easy for a university to move physically in the short run, the assumption of a fixed university also appears reasonable.

Figure 3.3 illustrates the location of the university and a particular student at $x < \frac{1}{2}$.

We formulate the utility function of a student, who is located at $x < \frac{1}{2}$, such as:

$$u = t - s - k\left(\frac{1}{2} - x\right). \tag{3.12}$$

Whether the mobility cost is linear or non-linear is critical when the university chooses its location, but it does not affect the nature of the analysis in this fixed location model. Equation (3.12) assumes that the tuition fee linearly affects the utility of students.

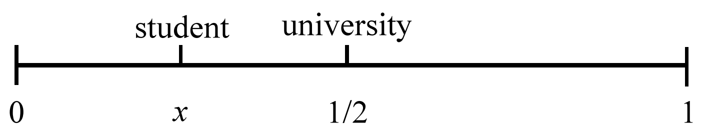


Figure 3.3. Location of a university and students.

Assuming without loss of generality that the outside option other than enrolling at the university in this jurisdiction gives each student zero utility ($u = 0$), we can find a particular $\hat{x}$ who is indifferent between enrolling and not. This condition satisfies $t - s - k\left(\frac{1}{2} - \hat{x}\right) = 0$, so that $\hat{x} = \frac{1}{2} + \frac{s-t}{k}$. It can be easily shown that because of the symmetric characteristics, $\hat{x} = \frac{1}{2} + \frac{t-s}{k}$ for students who are located at $x > \frac{1}{2}$. From these, student enrollment can be represented as $n = n(t, k, s) = \frac{2(t-s)}{k}$.

Since $n \in [0, 1]$ is assumed, t must be bounded such that $t \in [s, s + \frac{k}{2}]$. The condition, $t > s$, implies that the teaching value should be larger than the tuition fee for the university to obtain positive student enrollment. It is no wonder that students, who intend to attend the university, demand high-level teaching comparable to the tuition fee that they must pay. By contrast, if $t \leq s$, the university cannot obtain any students ($n = 0$). Also, even if teaching effort is excessive to the point that $t \geq s + \frac{k}{2}$, the intake of students cannot be higher than 1 ($n = 1$). As expected, $\frac{\partial n}{\partial t} = \frac{2}{k} > 0$, $\frac{\partial n}{\partial k} = -\frac{2(t-s)}{k^2} \leq 0$, $\frac{\partial n}{\partial s} = -\frac{2}{k} < 0$, $\frac{\partial^2 n}{\partial t} = 0$, $\frac{\partial n^2}{\partial t \partial k} = -\frac{1}{k^2} < 0$, and $\frac{\partial^2 n}{\partial t \partial s} = 0$ are confirmed, which also satisfies the previous assumptions.

Most of these functional forms regarding a university and students seem quite specific, but as presented above, they are made to closely reflect the mechanism in force and help us gain a better sense of complex interactions that exist between research and teaching activities. In line with this, we further the analysis hereafter.

The university maximizes its payoff given Stage 2 (the decision of students):

$$\max_{r,t} U(r,t) = rb - \left(\frac{r^2}{2} + \frac{t^2}{2} + \varepsilon rt\right) \tag{3.13}$$

$$\text{subject to } b = F + sn \text{ and } n = \frac{2(t-s)}{k} \in [0,1]. \tag{3.14}$$

From Equations (3.13) and (3.14), the payoff function of the university is abbreviated into $U(r,t) = r[F + \frac{2s(t-s)}{k}] - (\frac{r^2}{2} + \frac{t^2}{2} + \varepsilon rt)$. The first-order condition of maximizing $U(r,t)$ with regard to r and t is formulated as follows:

$$\frac{\partial U}{\partial r} = b - r - \varepsilon t = 0, \tag{3.15}$$

$$\frac{\partial U}{\partial t} = r\left(\frac{2s}{k}\right) - t - \varepsilon r = 0. \tag{3.16}$$

Let us define the solution of Equations (3.15) and (3.16) as $\hat{e} = (\hat{r}, \hat{t})$. (The “hat" denotes a solution of the simultaneous equations, but not necessarily an equilibrium solution.) Although the calculated result, $\hat{r}$ and $\hat{t}$, may be strictly negative, we have to keep in mind that research and teaching effort cannot be negative in principle. We are interested only in the case where the both types of effort are positive. One reason for this assumption may be supported by the idea that the financing agency usually provides a minimum research fund enough to enable a positive level of research and teaching activities. (See also Lemma 3.2 below.) On the other hand, we have to examine the second-order condition of the maximization problem. Generally, in order for the payoff function to have a global maximum, it must be a strictly concave function. For strict concavity, $(1-\varepsilon^2)k^2 + 4ks\varepsilon - 4s^2 < 0$, that is, $\varepsilon \in (-1+\frac{2s}{k}, 1+\frac{2s}{k})$ is required, which suggests that ε must be bounded both upward and downward. [15] Under a normal condition, it is expected that the degree of substitutability (complementarity) cannot diverge to infinity and thus falls within a finite range. The analysis proceeds assuring that this second-order condition is always satisfied.

Lemma 3.1 The solution, $\hat{e} = (\hat{r}, \hat{t})$, has the closed form such that:

$$\hat{r} = \frac{k(kF - 2s^2)}{(1-\varepsilon^2)k^2 + 4ks\varepsilon - 4s^2}, \tag{3.17}$$

$$\hat{t} = \frac{(2s - k\varepsilon)(kF - 2s^2)}{(1-\varepsilon^2)k^2 + 4ks\varepsilon - 4s^2}. \tag{3.18}$$

When both conditions, $F > \frac{2s^2}{k}$ and $\varepsilon \in (-1+\frac{2s}{k}, \frac{2s}{k})$, are satisfied, $\hat{e} = (\hat{r}, \hat{t})$ is a positive interior equilibrium solution with $\hat{r} > 0$ and $\hat{t} > 0$.

As we can see from Lemma 3.1, when conditions for both F and ε are satisfied, positive research and teaching effort is assured. The first condition is that a research

[15] The second derivatives of the function $U(r,t)$ are as follows : $\frac{\partial^2 U}{\partial r^2} = \frac{\partial^2 U}{\partial t^2} = -1$ and $\frac{\partial^2 U}{\partial r \partial t} = \frac{\partial^2 U}{\partial t \partial r} = \frac{2s}{k} - \varepsilon$. Hence, the second-order condition is calculated as $(\frac{\partial^2 U}{\partial r^2})(\frac{\partial^2 U}{\partial t^2}) - (\frac{\partial^2 U}{\partial r \partial t})^2 > 0$. This inequality expression can be simplified into $k^2\varepsilon^2 - 4ks\varepsilon + 4s^2 - k^2 < 0$. Solving it for ε, we obtain $-1+\frac{2s}{k} < \varepsilon < 1+\frac{2s}{k}$. Assuming the cost function, $C(r,t)$, is strictly concave, we can further confine the range of ε to $\varepsilon \in (-1+\frac{2s}{k}, 1)$.

fund is sufficiently large compared to a tuition fee (discounted by a mobility cost). This implies that unless a research fund is extremely small, the university can exert positive research effort. Moreover, the second condition for the degree of substitutability stipulates a bounded range that is narrower upward than that assumed before ($\varepsilon < 1 + \frac{2s}{k}$). [16]

Our interest largely lies in a positive interior solution where the university undertakes strictly positive research and teaching effort. With regard to this solution presented in Equations (3.17) and (3.18), does research and teaching effort increase as the degree of substitutability between these activities becomes small? At the same time, do research output and student enrollment increase, too, in tandem with the change in research and teaching effort? The answer is absolutely “yes", and they all increase. The following proposition describes this result.

Proposition 3.2 Consider a positive interior solution, $r^* = \hat{r} > 0$ and $t^* = \hat{t} > 0$. The smaller the degree of substitutability between research and teaching effort in terms of the cost function, the more effort the university dedicates to the both activities, and thereby, research output and student enrollment also increase. In other words, $\frac{\partial r^*}{\partial \varepsilon} < 0$, $\frac{\partial t^*}{\partial \varepsilon} < 0$, $\frac{\partial R^*}{\partial \varepsilon} < 0$, and $\frac{\partial n^*}{\partial \varepsilon} < 0$ hold. (The larger the degree of substitutability, the less effort the university dedicates to both activities.)

Proposition 3.1, which indicates that a small degree of substitutability operates in favor of research and teaching activities in an interior equilibrium solution is not a surprising result. However, the size of ε is important from the viewpoint of an actual

[16] When $\frac{2s}{k} > 1$ holds, we can rewrite the condition as $\varepsilon \in (-1 + \frac{2s}{k}, 1)$.

university policy and management. It seems that ε tends to be positive in most universities except some top-ranked general universities in which many high-achieving research students study. In fact, there are some well-organized universities that establish a favorable situation for both research and teaching activities. In such universities, better educated students help faculty members produce high-quality research output, for example, as co-authors. But realistically, they may be rare. Many university faculty members, especially in recent years, have been finding it more difficult to strike a fine balance between research and teaching activities, as demand for teaching responsibility grows. [17] Hence, in order to effectively conduct both research and teaching activities with limited resources, it is much more essential that policymakers or university administrative officials design institutional arrangements of universities towards reducing the degree of substitutability between these two activities. [18]

Supplementary note

Let us continue to consider a strictly positive interior solution, $r^* = \hat{r} = \frac{k(kF-2s^2)}{(1-\varepsilon^2)k^2+4ks\varepsilon-4s^2} > 0$ and $t^* = \hat{t} = \frac{(2s-k\varepsilon)(kF-2s^2)}{(1-\varepsilon^2)k^2+4ks\varepsilon-4s^2} > 0$. When it comes to the effect of a research fund, F, on research and teaching activities in this illustrative modeling, $\frac{\partial r^*}{\partial F} = \frac{k^2}{(1-\varepsilon^2)k^2+4ks\varepsilon-4s^2} > 0$

17 The tenure track system continues to apply pressure on university researchers who are not guaranteed lifetime employment. They are generally evaluated not only by outside experts with regard to research activities, but also by students of their belonged universities with regard to teaching activities. This evaluation system has been rapidly implemented at many universities in the world. Additionally, more emphasis has been placed on nurturing graduate students in hope of productive research output being attained by potentially quality researchers.

18 Demski and Zimmerman (2000), who succinctly investigate the question on “research versus teaching" in the academic community, acknowledge the fact that they could be substitutes in the short run, because the time academic staff can devote to research is limited due to teaching obligations. On the other hand, the authors also maintain that they could be mutually complementary activities in the long run, where research motivations of academic staff are frequently stirred by class notes, exams, students’ inquiries, and among others. They argue in their conclusion that academic staff should be encouraged to better exploit the opportunity to teach in order to generate more research output.

and $\frac{\partial t^*}{\partial F} = \frac{k(2s-k\varepsilon)}{(1-\varepsilon^2)k^2+4ks\varepsilon-4s^2} > 0$ are assured for $F > \frac{2s^2}{k}$ and $\varepsilon \in (-1+\frac{2s}{k}, \frac{2s}{k})$. By using the same demonstration with Proposition 3.1, we can also easily demonstrate that $\frac{\partial n^*}{\partial F} > 0$ and $\frac{\partial R^*}{\partial F} > 0$. This suggests that in the range of a positive interior solution, an increased research fund positively affects both research output and student enrollment, as long as the research fund is large and substitutability is not strong. It is noticeable that an extreme case does not appear, where an increased research fund reduces research output and student enrollment for the high degree of substitutability, as Proposition 3.1 (3) refers to this possibility. In this sense, this illustrative model indicates a normal research and teaching environment.

3.4.2. Analysis when the degree of substitutability is zero

The following investigation formulates an equilibrium solution as a specific case when the degree of substitutability between research and teaching activities is zero (i.e. $\varepsilon = 0$), which means that they are “independent". This simplification helps us to elicit precise effects of parameters change of a research fund and a tuition fee on university research and teaching activities. In later analyses, we see that the theoretical results are further extended as we add new assumptions and constraints compared to the basic results which were discussed previously.

A technical assumption is made about the allocation of a research fund F, contingent on teaching effort in order to avoid complexity of the analytical solution.

Assumption 3.1 The financing agency allocates no research funds ($F = 0$) to the university in the case where the university enrolls zero student ($n = 0$) through deficient teaching effort. [19]

This assumption is just technical, but intends to eliminate the case where the university can obtain a higher payoff by concentrating only on research activities, but not enrolling any new, prospective students. Since minimum student enrollment through teaching activities can be also viewed as an important mission in addition to research activities that most universities are required to fulfill, it is possible that the university does not qualify to receive any research funds without any students enrolling.

By substituting $\varepsilon = 0$ into Equations (3.17) and (3.18), we derive the following expressions:

$$\hat{r} = \frac{k(kF - 2s^2)}{k^2 - 4s^2}, \tag{3.19}$$

$$\hat{t} = \frac{2s(kF - 2s^2)}{k^2 - 4s^2}. \tag{3.20}$$

In order for the second-order condition to be satisfied, $k^2 - 4s^2 > 0$ needs to be assumed. By solving for s, we obtain $s < \frac{k}{2}$. (Note that unlike the previous exercise, the second-order condition cannot be solved for ε.) In addition, taking into account that $\hat{r}$ and $\hat{t}$ are usually positive values, we suppose that F is relatively large compared to s: that is, $F > \frac{2s^2}{k}$. [20] These two conditions imply that the tuition fee must not be extremely high, which appears to be supported by the fact that many countries

[19] A slightly different assumption can be made such that the financing agency allocates no research funds in the case where the university enrolls students less than $\underline{n} > 0$. Although this kind of alternatives does not change the intuition of the analysis, the simple assumption that has been already defined in the text is employed.

[20] All the units of F (research fund), s (tuition fee), and k (mobility cost) can be considered as any monetary unit, for example, US dollars. But it would not be practically appropriate to compare the amounts between them in a straightforward manner since the payoffs of a university and students are not always numerically comparable. Consequently, the magnitude relation among them should be regarded as relative.

carefully regulate it to keep it low in support of the students' welfare. Because of these conditions, a positive interior solution, $\hat{r} > 0$ and $\hat{t} > 0$, are satisfied as in Subsection 3.4.1.

According to the condition regarding student enrollment, $n \in [0, 1]$, we can derive the lemma that describes the equilibrium solutions, $\boldsymbol{e}^* = (r^*, t^*)$. (The "asterisk" denotes an equilibrium solutions as with the previous analyses.)

Lemma 3.2 Let us denote the closed form of $\hat{n}$ and $\hat{R}$ as:

$$\hat{n} = \frac{2s(2F - k)}{k^2 - 4s^2}, \tag{3.21}$$

$$\hat{R} = \hat{r}^2 = \left[\frac{k(kF - 2s^2)}{k^2 - 4s^2}\right]^2. \tag{3.22}$$

Suppose that $r^* + t^* < \bar{a}$ is satisfied at an equilibrium. (In other words, the capacity has "slack".) Then, the equilibrium solutions, $e^* = (r^*, t^*)$, n^*, and R^*, are provided by:

(1) **University closure**: $\boldsymbol{e}^* = (0, 0)$, $n^* = 0$, and $R^* = 0$ for $\frac{2s^2}{k} < F < s$;

(2) **Minimum teaching activities**: $\boldsymbol{e}^* = (F + \delta_r, s + \delta_t) \approx (F, s)$, $n^* = \delta_n \approx 0$, and $R^* = (F + \delta_r)^2 = F^2 + \delta_R \approx F^2$ for $s < F \leq \frac{k}{2}$;

(3) **Under-enrollment**: $\boldsymbol{e}^* = (\hat{r}, \hat{t})$, $n^* = \hat{n} \in (0, 1)$, and $R^* = \hat{R}$ for $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2 - 4s^2}{4s}$; and

(4) **Full enrollment**: $\boldsymbol{e}^* = (s + F, s + \frac{k}{2})$, $n^* = 1$, and $R^* = (s + F)^2$ for $F \geq \frac{k}{2} + \frac{k^2 - 4s^2}{4s}$, where δ_i with $i = r, t, n$, and R is an infinitesimal positive value.

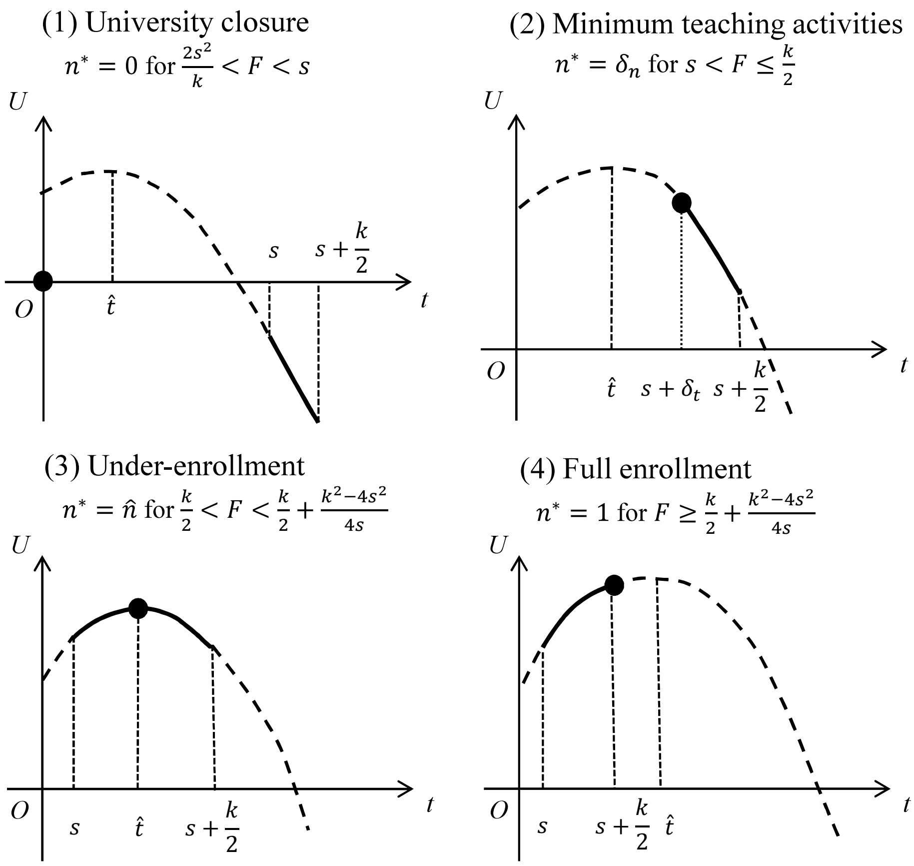


Figure 3.4. Equilibrium solutions with regard to t classified by F.

Lemma 3.2 demonstrates that the equilibrium solutions can differ because the range of student enrollment is confined to $n \in [0, 1]$. Figure 3.4 draws the configurations of how the equilibrium solution of teaching effort, t^*, changes in accordance with a research fund, F. The research fund is really used by research activities, but it also enhances an incentive of the university to conduct teaching activities since the rate of return in teaching activities becomes larger than research activities. (A more detailed explanation is provided in the next subsection.)

From Assumption 3.1, if the amount of a research fund is quite small, the university cannot afford to conduct any research and teaching activities ($r^* = 0$ and $t^* = 0$)

(Lemma 3.2 [1]: university closure) or conducts them only at minimum levels that assure slight student enrollment ($r^* = F + \delta_r$ and $t^* = s + \delta_t$) (Lemma 3.2 [2]: minimum teaching activities). In these two cases, the university cannot attain a significant amount of student enrollment. More notably, because research output is discontinuous between zero and F^2, the amount of a research fund distributed by the financing agency is critical for the university to engage in research activities. When producing more research output as well as acquiring significantly positive student enrollment, a certain amount of a research fund needs to be allocated to the university (Lemma 3.2 [3]: under-enrollment). Although a significantly positive number of students attend the university, under-enrollment ($n^* \in (0, 1)$) actually happens. By contrast, even if the university obtains a much larger monetary resource, it cannot increase student enrollment more than 1, but just $n^* = 1$ (Lemma 3.2 [4]: full enrollment).

3.4.3. Comparative statics in a simplified setting

We analyze comparative statics in the equilibrium solutions, especially when the degree of substitutability is zero (i.e. $\varepsilon = 0$). This postulate helps us elicit the impact of the parameters, in particular, a research fund on research and teaching activities.

Non-binding capacity constraint

Let us consider the case where the capacity constraint is not binding (i.e. $r^* + t^* < \overline{a}$). Namely, the university has a certain affordable capacity to commit effort for research and teaching. The following proposition answers how a research fund, F, affects the equilibrium solutions, $e^* = (r^*, t^*)$, n^*, and R^*, which are defined in Lemma 3.2.

Proposition 3.3 Suppose that $r^* + t^* < \overline{a}$ at the equilibrium solutions defined in Lemma 3.2. The comparative statics with regard to a research fund, F, indicates:

(1) With regard to research effort and research output:

(1-i) $\frac{\partial r^*}{\partial F} > 0$ and $\frac{\partial R^*}{\partial F} > 0$ for $F > s$; and

(1-ii) $\frac{\partial r^*}{\partial F} = \frac{\partial R^*}{\partial F} = 0$ for $\frac{2s^2}{k} < F < s$.

(2) With regard to teaching effort and student enrollment:

(2-i) $\frac{\partial t^*}{\partial F} > 0$ and $\frac{\partial n^*}{\partial F} > 0$ for $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2-4s^2}{4s}$; and

(2-ii) $\frac{\partial t^*}{\partial F} = \frac{\partial n^*}{\partial F} = 0$ for $\frac{2s^2}{k} < F \leq \frac{k}{2}$ and $F \geq \frac{k}{2} + \frac{k^2-4s^2}{4s}$.

(3) With regard to a change in F, $\frac{dr^*}{dt^*} > 1$ for $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2-4s^2}{4s}$.

It is clear that in the absence of substitutability, an increase in a research fund can boost research output except for the case where the university entirely shuts down its research and teaching activities (Proposition 3.3 [1-ii]). In particular, at the equilibrium of $\boldsymbol{e}^* = (\hat{r}, \hat{t})$, Proposition 3.1 (2-i) can immediately lead to $\frac{\partial R^*}{\partial F} > 0$ by setting $\frac{\partial^2 C}{\partial r \partial t} = 0$. This simple result may provide the support for the prevailing claim that devoting more resources into universities can stimulate research activities. Most recent empirical studies have observed that there is a positive correlation between R&D investment financed from the outside and research output in universities, and therefore, the result endorses a frequently observed common finding. [21] However, notice that this

[21] Payne and Siow (2003), who use a database of 18 U.S. research universities, find that an increase of 1 million U.S. dollar in federal research funding to a university generates 10 more articles and 0.2 more patents. Based on this observation, the authors argue that increasing research funds generates more research output, although it may not necessarily result in higher quality. With a particular focus on the Canadian nanotechnology field, Beaudry and Allaoui (2012) conclude that a greater amount of public funds certainly produce more research output of individual academics as represented by the number of scientific articles. Furthermore, Yonetani et al. (2013) investigate the relationship between the number of researchers and R&D expenditure as input and the number of articles as output in the field of natural science, using the panel data of *Web of Science and Survey of R&D* compiled for Japanese universities. They find that intramural expenditure of R&D funds received from external sources has a

result is highly dependent on the specific assumption of the absence of substitutability. As we already examined in Proposition 3.1, the high degree of substitutability could cause a decrease in research output although such situation would not be considered prevailing across universities.

Proposition 3.1 (2-i) also suggests that a research fund increases teaching activities as well in the case of under-enrollment (Proposition 3.3 [2-i]). As explained before, the intuition is as follows. At the beginning, the university can afford to devote more effort to research activities owing to an increased research fund. At the same time, it becomes more profitable to dedicate some effort toward teaching activities than research activities because the marginal payoff obtained from research activities has dropped. In the next stage, since the utilities of students are fostered by the improved teaching effort, student enrollment is also expected to increase, and thereby, marginal students decide to enroll at the university. A substantial increase in student enrollment contributes to enriching the budget of the university through tuition revenue. This is how research and teaching activities indirectly interact with each other not relying on their substitutability, and naturally increasing their amounts in the process. In short, the “multiplier effect" is in force between research and teaching effort in response to an increase in the research fund.

This intuition is much easier to understand by looking at Figure 3.5 that depicts the response functions in the diagram of (r, t): $r(t) = (\frac{2s}{k})t + \frac{kF-2s^2}{k}$ and $t(r) = (\frac{2s}{k})r$. The intersection of the two lines denoted by point A represents an initial equilibrium

positive correlation with articles published by researchers at both national and private universities.

solution (r^*, t^*). Note that when F increases, $r(t)$ shifts outward (right-hand side). If a research fund is increased and teaching effort is kept constant at the level of t^*, the combination of research and teaching effort moves to point B, and r^* also increases to $\check{r}$. But because t^* is no longer an optimal at point B, the equilibrium solution ends up at point C, where further increases in both research and teaching effort occur, (r^{**}, t^{**}). Moreover, since the slope of $r(t)$ $(t(r))$ is larger (smaller) than 1, we can see that the more increased effort is diverted to research rather than teaching (Proposition 3.3 [3]).

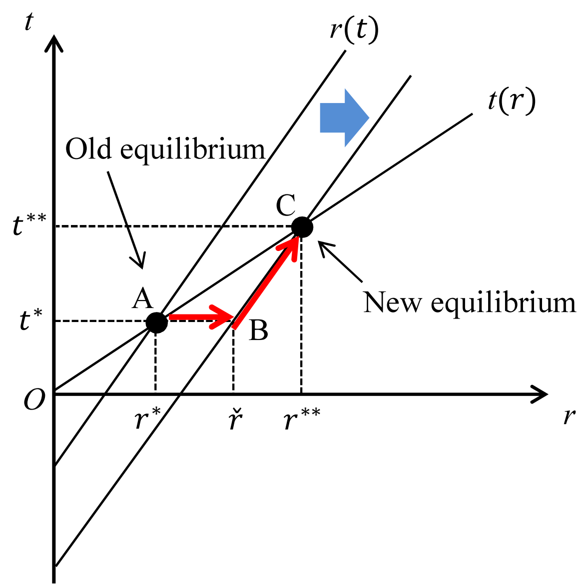


Figure 3.5. Effect of an increase in F on research and teaching effort.

Continuously focusing on the interior equilibrium solution (that is, non-binding capacity case), $\boldsymbol{e}^* = (\hat{r}, \hat{t})$, we derive the comparative statics regarding k (mobility cost) and s (tuition fee).

Proposition 3.4 Consider the interior equilibrium solutions, $\boldsymbol{e}^* = (\hat{r}, \hat{t})$, $n^* = \hat{n}$, and $R^* = \hat{R}$. The comparative statics with regard to k and s indicates:

(1) $\frac{\partial r^*}{\partial k} < 0$, $\frac{\partial t^*}{\partial k} < 0$, $\frac{\partial n^*}{\partial k} < 0$, and $\frac{\partial R^*}{\partial k} < 0$; and

(2) $\frac{\partial r^*}{\partial s} > 0$, $\frac{\partial t^*}{\partial s} > 0$, $\frac{\partial n^*}{\partial s} > 0$, and $\frac{\partial R^*}{\partial s} > 0$.

What is noteworthy is that the effects of a mobility cost and a tuition fee operate in a different direction in this specific illustrative case, although they are similar in that both of them lower the utilities of students. More precisely, whereas a rise in a mobility cost causes a reduction in university activities concerning both research and teaching, a rise in a tuition fee in contrast gives the university an incentive to improve these two activities.

In fact, a higher mobility cost definitely decreases student enrollment, and thus, reduces the budget of the university at that rate, which also culminates in a decrease in research output. Like the mechanism that works in a mobility cost, a rise in a tuition fee actually decreases student enrollment at an initial stage. Nevertheless, the university may still be able to increase tuition revenue as a whole. The university is incentivized to make more teaching effort by a rise in the tuition fee, as it can earn more tuition revenue per student. In association with such increased teaching effort, the contribution to tuition revenue from an intramarginal population of students is large as compared to the loss from marginal students who do not apply. This result stems from the fact that the student enrollment function, $\hat{n}$, defined at the equilibrium is “inelastic" with regard to a tuition fee in our illustrative model. [22] In turn, this positive effect on the research budget can strengthen the incentive of the university for both research and teaching effort. Accordingly, in spite of the negative effect on

[22] In checking the tuition fee elasticity of the student enrollment function, $\hat{n} = \frac{2s(2F-k)}{k^2-4s^2}$, it is sufficient to examine the sign of $\hat{n} + s(\frac{\partial \hat{n}}{\partial s})$; if $\hat{n} + s(\frac{\partial \hat{n}}{\partial s}) < 0$ (> 0), then $\hat{n}$ is elastic (inelastic). We obtain $\hat{n} + s(\frac{\partial \hat{n}}{\partial s}) = \frac{2s(2F-k)}{k^2-4s^2} + s[\frac{2(2F-k)(k^2-4s^2)+16s^2(2F-k)}{(k^2-4s^2)^2}] = \frac{4k^2 s(2F-k)}{(k^2-4s^2)^2} > 0$ under the presumed assumption of $F > \frac{k}{2}$. Hence, the student enrollment function, $n^* = \hat{n}$, is inelastic for a tuition fee.

student enrollment, the university is expected to achieve higher research output and student enrollment than prior to a rise in a tuition fee. [23]

From the above-mentioned result, we may be tempted to arrive at a hasty conclusion that the higher we set a tuition fee, the more we can expect research output and student enrollment to increase. But this is not always true for the reason that the tuition fee, s, is restricted by the condition, $F > \frac{2s^2}{k}$ and $s < \frac{k}{2}$, which requires that the tuition fee must be kept low enough. It is therefore impossible and undesirable to arbitrarily raise a tuition fee to increase research and teaching activities.

Supplementary note

One may have a doubt as to whether these results of comparative statics are robust in a more general setting because they are derived from a particular parameter setting. Subsection 3.8.3, which analyzes a general model as defined in Section 3.3, illustrates that the signs of comparative statics can be changed by substitutability between research and teaching activities. If we present an assumption that the degree of substitutability is zero, it can be demonstrated that a rise in the mobility cost, k, negatively affects research and teaching activities. On the other hand, it is revealed that the effect of a rise in the tuition fee is not necessarily decisive depending on other parameters, even when the degree of substitutability is zero.

Another possible doubt might be that the result of comparative statics regarding a tuition fee is different from the reality: student enrollment is highly likely to decrease

[23] In the process where research effort is enhanced based on the more enriched budget, the “multiplier effect" is also operating between research and teaching effort.

in accordance with a rise in a tuition fee under the competition between universities. But since the current model assumes a single university inside the jurisdiction, it boldly eliminates the competition effect of a reduction in a tuition fee. Competition for students will be mentioned in Section 3.5 that posits multiple universities.

Binding capacity constraint

Next, let us consider the case where the university fully exerts its capacity, that is, the capacity constraint is “binding" (i.e. $\hat{r} + \hat{t} > \overline{a}$). From hereupon, we mainly focus on under-enrollment, $n^* \in (0, 1)$, described by Lemma 3.2 (3), which seems the most common in real university-student markets. The following proposition points out a clear-cut opposite conclusion from Proposition 3.3 as to the effect on teaching effort and student enrollment.

Proposition 3.5 Suppose that the capacity constraint of the university is binding, and in other words, $\hat{r} + \hat{t} > \overline{a}$ holds. At an equilibrium solution $e^* = (r^*, t^*)$ with $r^* + t^* = \overline{a}$, we obtain $\frac{\partial r^*}{\partial F} > 0$, $\frac{\partial t^*}{\partial F} < 0$, $\frac{\partial n^*}{\partial F} < 0$, and $\frac{\partial R^*}{\partial F} > 0$.

The mechanism behind this proposition is quite straightforward. Enriching the budget enables the university to engage in more research activities. Nevertheless, since the effort has already reached a maximum level, teaching effort is in turn reduced, and as a result, student enrollment is certain to diminish. When the capacity is fully exerted, an increased research fund allocated by a financing agency ends up “crowding out" teaching activities, and thereby, student enrollment too. [24] As sharply contrasted

[24] When research and teaching activities are independent, a decrease in tuition revenue caused by reduced student enrollment has a relatively small impact on the total research budget, compared to an

with the previous “multiplier effect", we can name this polar change the “crowding-out effect". From this, we can see that if a capacity constraint is introduced into the model, a decrease in teaching effort could be caused even in the absence of substitutability between research and teaching activities.

As we observed in the empirical implications suggested in Section 3.2, it is probable that a decrease in doctorates awarded and student enrollment for private universities might be caused by an increase in a research fund for the reason of the high degree of substitutability or crowding-out effect. But it is empirically difficult to discern these two causes from our limited dataset.

3.5. Tuition fee is a control variable

Now suppose that a tuition fee is no longer exogenously fixed, but an endogenously controlled variable that is set to maximize tuition revenue. It could be that a government authority intends to control tuition fees of universities since, in doing so, it may be able to save on research spending that is distributed to universities. [25] Or it could also be that universities are allowed to freely determine their tuition fees in order to maximize their payoffs. As explained below, these two interpretations are mathematically equivalent from an analytical viewpoint.

increase in the research fund. This is because the university maximizes its payoff ultimately through the optimal choice of research output, and thus, intends to maintain the total research budget. See also the demonstration of Proposition 3.5 in Subsection 3.8.1.

[25] In addition to research funds, U.S. universities have been required to earn their own research resources through patents and consulting revenue especially since the enactment of the Bayh-Dole Act in the U.S. As regards to the studies on revenue earning activities by universities, see Mowery and Ziedonis (2002), Azuoray et al. (2009), and Jensen et al. (2010).

The timing of the model (presented in Subsection 3.3.2) is slightly modified to include a government authority's decision in Stage 1.5; the government authority determines the tuition fee, s, of the university between Stages 1 and 2.

3.5.1. Analysis of a general case

In the first place, what is considered is the government authority problem of finding an optimal tuition fee, s^*, that maximizes university tuition revenue. Letting E denote this revenue, we define the maximization problem such that:

$$\max_s E = sn(t, k, s). \tag{3.23}$$

On this problem, the first-order condition for s is rendered by:

$$\frac{\partial E}{\partial s} = n + s\left(\frac{\partial n}{\partial s}\right) = 0. \tag{3.24}$$

Solving Equation (3.24) by s, we can obtain $s = s(t; k)$ as a function of t (and k). If it is assumed that the university is allowed to choose an optimal tuition fee by itself, the first-order condition of maximizing $U = r[F + sn(t, k, s)] - C(r, t)$ is given by $r[n + s(\frac{\partial n}{\partial s})] = 0$. By positing $r > 0$, we can find the same condition as the above.

In order to check whether the solution has a global maximum, we derive the second-order condition: $\frac{\partial^2 E}{\partial s^2} = 2(\frac{\partial n}{\partial s}) + s(\frac{\partial^2 n}{\partial s^2}) < 0$. We can see that unless $n(t, k, s)$ is a strong convex function against s (i.e. $\frac{\partial^2 n}{\partial s^2} > 0$), this condition is not violated. But we hereafter proceed by assuming that $\frac{\partial^2 E}{\partial s^2} < 0$ is satisfied at $s = s(t; k)$. (In an illustrative case discussed in Section 3.4, since $n = \frac{2(t-s)}{k}$ holds, the second-order condition is always satisfied.)

If we take a derivative on both sides of Equation (3.24) by t, we obtain:

$$\begin{aligned}
&\frac{\partial n}{\partial t} + 2\left(\frac{\partial n}{\partial s}\right)\left(\frac{\partial s}{\partial t}\right) + s\left(\frac{\partial^2 n}{\partial s \partial t}\right) + s\left(\frac{\partial^2 n}{\partial s \partial t}\right) = 0. \\
&\Longleftrightarrow \underbrace{\left[2\left(\frac{\partial n}{\partial s}\right) + s\left(\frac{\partial^2 n}{\partial s^2}\right)\right]}_{\text{negative}}\left(\frac{\partial s}{\partial t}\right) = -\left[\frac{\partial n}{\partial t} + s\left(\frac{\partial^2 n}{\partial t \partial s}\right)\right].
\end{aligned} \tag{3.25}$$

If we also suppose $\frac{\partial^2 n}{\partial t \partial s}$=0, that is, there exist no cross terms between t and s in the function of $n(t, k, s)$, we can derive $\frac{\partial s}{\partial t} > 0$ from the assumption.

By using an optimal tuition fee, $s = s(t; k)$, we redefine the student enrollment function as $n(t, k, s) = n(t, k, s(t; k)) = \tilde{n}(t; k)$. Based on these settings, we confirm in what follows the Hessian matrix of $U(r, t) = r[F + s(t; k)\tilde{n}(t; k)] - C(r, t)$, the payoff function of a university. At the beginning, let us consider the first-order condition for maximizing $U(r, t)$ (the “asterisk" is omitted from the following expressions for descriptive simplicity):

$$\frac{\partial U}{\partial r} = F + s\tilde{n} - \frac{\partial C}{\partial r} = 0, \tag{3.26}$$

$$\frac{\partial U}{\partial t} = r\left[\tilde{n}\left(\frac{\partial s}{\partial t}\right) + s\left(\frac{\partial \tilde{n}}{\partial t}\right)\right] - \frac{\partial C}{\partial t} = 0. \tag{3.27}$$

From Equations (3.26) and (3.27), we can find an equilibrium solution, $\boldsymbol{e}^* = (r^*, t^*)$, and an optimal tuition fee, $s^* = s(t^*; k)$.

The second derivatives of $U(r, t)$ are provided by:

$$\frac{\partial^2 U}{\partial r^2} = -\frac{\partial^2 C}{\partial r^2} < 0, \tag{3.28}$$

$$\frac{\partial^2 U}{\partial t^2} = r\left[\tilde{n}\left(\frac{\partial^2 s}{\partial t^2}\right) + s\left(\frac{\partial^2 \tilde{n}}{\partial t^2}\right) + 2\left(\frac{\partial s}{\partial t}\right)\left(\frac{\partial \tilde{n}}{\partial t}\right)\right] - \frac{\partial^2 C}{\partial t^2}, \tag{3.29}$$

$$\frac{\partial^2 U}{\partial r \partial t} = \tilde{n}\left(\frac{\partial s}{\partial t}\right) + s\left(\frac{\partial \tilde{n}}{\partial t}\right) - \frac{\partial^2 C}{\partial r \partial t}. \tag{3.30}$$

Hence, the Hessian matrix, $\tilde{U} = \begin{bmatrix} \frac{\partial^2 U}{\partial r^2} & \frac{\partial^2 U}{\partial r \partial t} \\ \frac{\partial^2 U}{\partial r \partial t} & \frac{\partial^2 U}{\partial t^2} \end{bmatrix}$, is specified as:

$$\tilde{U} = \begin{bmatrix} -\frac{\partial^2 C}{\partial r^2} & \tilde{n}\left(\frac{\partial s}{\partial t}\right) + s\left(\frac{\partial \tilde{n}}{\partial t}\right) - \frac{\partial^2 C}{\partial r \partial t} \\ \tilde{n}\left(\frac{\partial s}{\partial t}\right) + s\left(\frac{\partial \tilde{n}}{\partial t}\right) - \frac{\partial^2 C}{\partial r \partial t} & r\left[\tilde{n}\left(\frac{\partial^2 s}{\partial t^2}\right) + s\left(\frac{\partial^2 \tilde{n}}{\partial t^2}\right) + 2\left(\frac{\partial s}{\partial t}\right)\left(\frac{\partial \tilde{n}}{\partial t}\right)\right] - \frac{\partial^2 C}{\partial t^2} \end{bmatrix}. \tag{3.31}$$

We obtain the determinant of $\tilde{U}$ as follows:

$$\begin{aligned} |\tilde{U}| = {} & \underbrace{\frac{\partial^2 C}{\partial r^2}}_{\text{positive}} \left[\frac{\partial^2 C}{\partial t^2} - r\left[\tilde{n}\left(\frac{\partial^2 s}{\partial t^2}\right) + s\left(\frac{\partial^2 \tilde{n}}{\partial t^2}\right) + 2\left(\frac{\partial s}{\partial t}\right)\left(\frac{\partial \tilde{n}}{\partial t}\right)\right]\right] \\ & - \underbrace{\left[\tilde{n}\left(\frac{\partial s}{\partial t}\right) + s\left(\frac{\partial \tilde{n}}{\partial t}\right) - \frac{\partial^2 C}{\partial r \partial t}\right]^2}_{\text{positive}}. \end{aligned} \tag{3.32}$$

It is necessary that $\frac{\partial^2 U}{\partial t^2} = r\left[\tilde{n}\left(\frac{\partial^2 s}{\partial t^2}\right) + s\left(\frac{\partial^2 \tilde{n}}{\partial t^2}\right) + 2\left(\frac{\partial s}{\partial t}\right)\left(\frac{\partial \tilde{n}}{\partial t}\right)\right] - \frac{\partial^2 C}{\partial t^2} < 0$ (Equation [3.29]) holds for $U(r, t)$ to indicate a global maximum, but in turn, the sign of $|\tilde{U}|$ in Equation (3.32) is indecisive. (As the degree of substitutability, $\frac{\partial^2 C}{\partial r \partial t}$, becomes small, it is much more difficult to satisfy $|\tilde{U}| > 0$.) Thus, while $U(r, t)$ has a global maximum at $\boldsymbol{e}^* = (r^*, t^*)$ and $s^* = s(t^*; k)$ for $|\tilde{U}| > 0$, a saddle point emerges for $|\tilde{U}| < 0$.

3.5.2. Analysis of an illustrative case

Let us revert to the illustrative case formulated in Section 3.4 to derive explicit equilibrium solutions. We continue to assume $\frac{\partial^2 C}{\partial r \partial t} = \varepsilon = 0$ for analytical simplicity. The maximization problem of university tuition revenue is defined such that:

$$\max_s E = sn = \frac{2s(t-s)}{k}. \tag{3.33}$$

The optimal solution is $s = \frac{t}{2}$, which suggests that the optimal tuition fee is one half of the teaching value. From Equation (3.33), the maximum value of E is $E = \frac{t^2}{2k}$.

Substituting the optimal s back into the university payoff function, we obtain:

$$U(r,t) = r\left(F + \frac{t^2}{2k}\right) - \left(\frac{r^2}{2} + \frac{t^2}{2}\right) = Fr - \frac{r^2}{2} + \left(\frac{r-k}{2k}\right)t^2. \tag{3.34}$$

The first-order conditions with regard to r and t of Equation (3.34) are provided by:

$$\frac{\partial U}{\partial r} = F - r + \frac{t^2}{2k} = 0, \tag{3.35}$$

$$\frac{\partial U}{\partial t} = \left(\frac{r-k}{k}\right)t = 0. \tag{3.36}$$

Equations (3.35) and (3.36) provide two possible values that induce $\frac{\partial U}{\partial r} = \frac{\partial U}{\partial t} = 0$: that is, $(r,t) = (F,0)$ and $(k, \sqrt{2k(k-F)})$ for $F < k$. These two points are depicted as Points A and S, respectively, in Figure 3.6. Let us focus on the point, $\tilde{e} = (\tilde{r}, \tilde{t}) = (k, \sqrt{2k(k-F)})$ assuming $F < k$. By solving Equation (3.35) with regard to t, we derive $t = r^{-1}(t) = \sqrt{2k(r-F)}$ with $r > F$. Focusing on the term $(\frac{r-k}{2k})t^2$ of Equation (3.34), we see that the larger (smaller) the t is for $r > k$ $(r < k)$, the higher payoff the university gains, and t is irrelevant to the payoff for $r = k$. This implies that $\tilde{e} = (k, \sqrt{2k(k-F)})$ with $F < k$ is a saddle point. In the range of $F < k$, we can also see that $(r,t) = (F,0)$ achieves a local maximum. Moreover, when $F > k$ holds, $(r,t) = (F,0)$ becomes a saddle point. (The point $\tilde{e} = (k, \sqrt{2k(k-F)})$ disappears for

$F > k$.) Subsection 3.8.2 demonstrates these arguments by examining the Hessian matrix of $U(r,t)$ as a quadratic approximation.

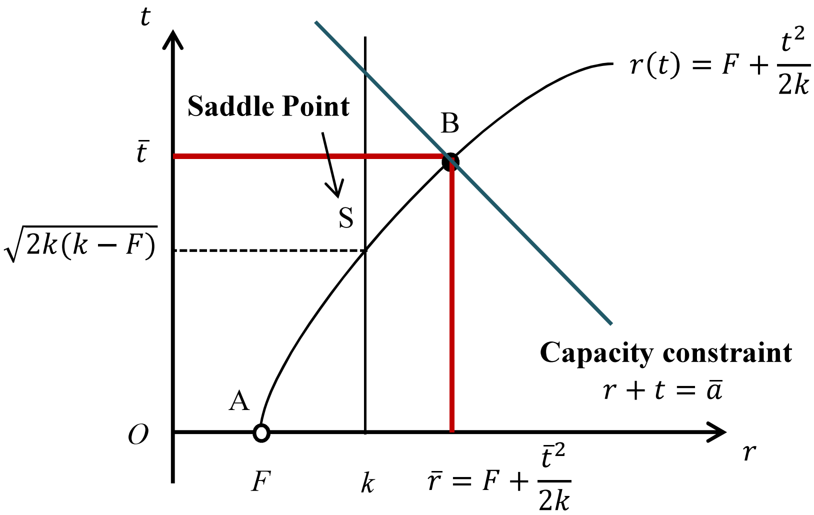


Figure 3.6. Reaction curve of $r(t)$ and the capacity constraint.

We make a following assumption of research activities conducted by a university.

Assumption 3.2 A university is required to choose an optimal amount of research effort for any amount of given teaching effort.

Assumption 3.2 implies that Equation (3.35), $r(t) = F + \frac{t^2}{2k}$, applies to any $t > 0$, which is critical for the equilibrium solution to be derived in Proposition 3.6 when the capacity constraint of a university is bounded. This assumption is analytically needed to explicitly identify an equilibrium solution, and to eliminate the case where the university obtains a higher payoff by decreasing research but increasing teaching effort in the left region from $r(t) = F + \frac{t^2}{2k}$. [26] It is suggested that whatever research output is

[26] Taking a total differential of Equation (3.34) provides $dU = (F + \frac{t^2}{2k})dr + (\frac{rt}{k})dt - (rdr + tdt) = (F + \frac{t^2}{2k} - r)dr + [\frac{(r-k)t}{k}]dt$, where $dr < 0$ and $dt > 0$. If we evaluate the effect of a minute change (first-order approximation) in r and t along the line of $r(t) = F + \frac{t^2}{2k}$, we obtain $dU = [\frac{(r-k)t}{k}]dt > 0$ for $r > k$ because $F + \frac{t^2}{2k} - r = 0$. The university can therefore increase its payoff by marginally decreasing r with the capacity constraint being binded.

potentially produced, a university may not be allowed to make light of the responsibility that comes with research dedication. This postulate could be justifiable on ground that a government authority sometimes intends to maintain the level of research effort conducted by a university, or that faculties of a university are reluctant to be swayed by university administrative officials into being forced to decrease research effort. In other words, it is assumed that a university cannot flexibly change research effort, and thus, determines it so as to maximize the payoff in accordance with each level of teaching effort. This paper is for probing implicative results, so that a complete investigation of implicit possible equilibria is not the ultimate goal. For this reason, Assumption 3.2 is made in later analyses.

Recall again that research and teaching effort of the university is bounded by its capacity, namely, $r + t \leq \overline{a}$. Hence, a corner solution, $\overline{e} = (\overline{r}, \overline{t}) > 0$, which satisfies both $r + t = \overline{a}$ and $r = F + \frac{t^2}{2k}$, can be an equilibrium solution that achieves a maximum university payoff, since more active research and teaching effort can generate a higher payoff for the university. Calculating these two simultaneous equations, we obtain $\overline{r} = \overline{a} + k - \sqrt{k(k + 2\overline{a} - 2F)}$ and $\overline{t} = -k + \sqrt{k(k + 2\overline{a} - 2F)}$ (where $k + 2\overline{a} - 2F > 0 \Leftrightarrow \overline{a} > F - \frac{k}{2}$ is assumed). Hereafter, $\overline{t}$ is conveniently used instead of $\overline{a}$ to denote a corner solution for descriptive simplicity.

Given that $\overline{e} = (\overline{r}, \overline{t}) = (F + \frac{\overline{t}^2}{2k}, \overline{t})$ is an equilibrium solution, we can also represent $\overline{s} = \frac{\overline{t}}{2}$, $\overline{n} = \frac{2(\overline{t} - \overline{s})}{k} = \frac{\overline{t}}{k}$, and $\overline{R} = \overline{r}(F + \overline{s}\,\overline{n}) = (F + \frac{\overline{t}^2}{2k})^2$, respectively. Additionally, we continue to assume that even if the full capacity is attained, under-enrollment in the jurisdiction, $n^* \in (0, 1)$, still exists – this implies that the number of prospective students who intend to attend the university is huge. Based on these derivations, we lead to Proposition 3.6 that describes equilibrium solutions.

Proposition 3.6 Suppose that under-enrollment, $n^* \in (0, 1)$, occurs. In the case where a tuition fee is a control variable, we obtain the equilibrium solutions, $\boldsymbol{e}^* = (r^*, t^*)$, n^*, s^*, and R^* as follows:

(1) With regard to $F < k$:

(1-i) **Large university**: $\boldsymbol{e}^* = \bar{\boldsymbol{e}} = (F + \frac{\bar{t}^2}{2k}, \bar{t})$, $s^* = \bar{s} = \frac{\bar{t}}{2}$, $n^* = \bar{n} = \frac{\bar{t}}{k}$, and $R^* = \overline{R} = (F + \frac{\bar{t}^2}{2k})^2$ for $\bar{t} > 2\sqrt{k(k-F)}$; and

(1-ii) **Small research institute (or small college)**: $\boldsymbol{e}^* = \boldsymbol{e}^0 = (F + \delta_r, \delta_t) \approx (F, 0)$, $s^* = s^0 \approx 0$, $n^* = n^0 \approx 0$, and $R^* = R^0 \approx F^2$ for $\bar{t} < 2\sqrt{k(k-F)}$,

(2) With regard to $F > k$, $\boldsymbol{e}^* = \bar{\boldsymbol{e}} = (F + \frac{\bar{t}^2}{2k}, \bar{t})$, $s^* = \bar{s} = \frac{\bar{t}}{2}$, $n^* = \bar{n} = \frac{\bar{t}}{k}$, and $R^* = \overline{R} = (F + \frac{\bar{t}^2}{2k})^2$ for every $\bar{t}$.

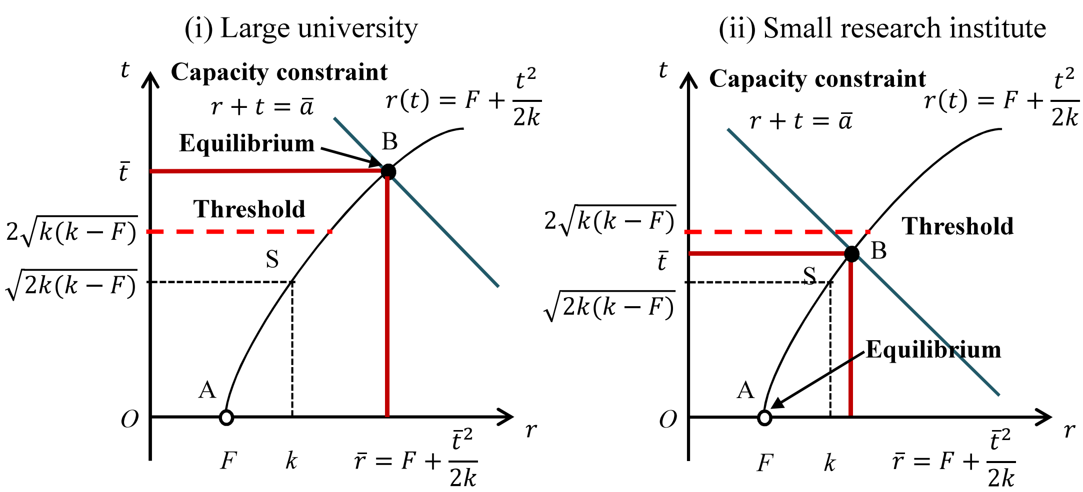


Figure 3.7. Equilibrium of a large university and a small research institute.

Figure 3.7 (i) and (ii) illustrate in diagram (r, t) the two polar equilibrium solutions of “large university" and “small research institute" demonstrated by Proposition 3.6

(1-i) and (1-ii), respectively. Under the assumption of $F < k$, [27] research and teaching effort is made by using a maximum capacity, $\overline{a}(= \overline{r}+\overline{t})$, associated with the equilibrium solution, $\boldsymbol{e}^* = \overline{\boldsymbol{e}} = (F+\frac{\overline{t}^2}{2k}, \overline{t})$, only if the capacity is sufficiently large ($\overline{t} > 2\sqrt{k(k-F)}$). To put it simply, this pattern is the case with a “large university" that can afford to get involved in a number of activities. In this case, Point B in Figure 3.7 (i) indicates the equilibrium solution. However, there is another possibility that if a so-called “small research institute" with a small capacity ($\overline{t} < 2\sqrt{k(k-F)}$) operates in the jurisdiction, the university will choose a minimum combination of research and teaching effort approximated by $\boldsymbol{e}^* = \boldsymbol{e}^0 = (F + \delta_r, \delta_t) \approx (F, 0)$, as shown at Point A in Figure 3.7 (ii). The infinitesimally small teaching efforts, δ_t, at the equilibrium may seem a bit extreme. But if Assumption 3.1 regarding minimum student enrollment (i.e. more than just zero) is modified as pointed out in Footnote 20, we can derive an equilibrium teaching effort that is of significantly positive value, $\underline{t} \in (\delta_t, \overline{t})$, but not an infinitesimal one. In such an interpretation, the university may as well be termed a “small college".

Why does a university prefer to conduct a minimum amount of activities? The reason is intuitively explained as follows: if the potential capacity is small enough, the university finds it difficult to benefit from “economies of scale" in research and teaching effort. An increase in effort reduces the payoff from the beginning up until the saddle point, $\tilde{\mathbf{e}} = (k, \sqrt{2k(k-F)})$ as shown in Point S, but in turn, they are likely to begin to improve the payoff past the point since the university can impose a higher

[27] One may think that the condition $F < k$ seems odd since a research fund is normally greater than a mobility cost per student. But aside from the explanation provided in Footnote 21, the possible reason why both F and k are comparable is that they can be measured in the same unit, namely, the total number of students is calculated as 1.

tuition fee on more present as well as incoming students. In such a situation, the “small research institute" with a small capacity cannot benefit from exploiting its capacity to the fullest before it reaches a point over which increased effort can provide a higher payoff to the university.

On the other hand, when the research fund is large enough to satisfy $F > k$, the payoff of the university becomes larger as effort increases along with the function, $r = r(t) = F + \frac{t^2}{2k}$. As a result, since the saddle point appears at $(F, 0)$ as exhibited in Point A, exerting its effort to full capacity is always optimal for the university. As is also easily demonstrated, the assumption of $F > k$ ensures that all the payoffs located on $r = r(t)$ are always strictly positive for the reason that the minimum payoff at $\boldsymbol{e}^0 = (F + \delta_r, \delta_t)$ is given by $U(F + \delta_r, \delta_t) \approx \frac{F^2}{2} > 0$.

The important point being made here is that even being in an ideal position to be a monopolist over student enrollment, a “small research institute", which cannot afford a large scale of activities, may place its smallest amount of research and teaching effort far below its potential capacity. Consequently, the strong support extended toward making use of research funds by a government authority can be justified especially for a “small research institute", in order to make the university choose an effort level that exploits its full capacity, and thereby, to shift to an equilibrium creating higher research output and student enrollment.

The following proposition regarding comparative statics with regard to a research fund exhibits a contrasting result to that was shown in Proposition 3.3.

Proposition 3.7 The comparative statics with regard to F for the equilibrium solutions led by Proposition 3.6 indicates that:

(1) $\frac{\partial r^*}{\partial F} > 0$, $\frac{\partial t^*}{\partial F} < 0$, $\frac{\partial s^*}{\partial F} < 0$, $\frac{\partial n^*}{\partial F} < 0$, and $\frac{\partial R^*}{\partial F} > 0$ for Proposition 3.6 (1-i) and (2); and

(2) $\frac{\partial r^*}{\partial F} > 0$, $\frac{\partial t^*}{\partial F} = 0$, $\frac{\partial s^*}{\partial F} = 0$, $\frac{\partial n^*}{\partial F} = 0$, and $\frac{\partial R^*}{\partial F} > 0$ for Proposition 3.6 (1-ii).

The result of Proposition 3.7 (1) that an increase in a research fund crowds out teaching activities is the same mechanism as Proposition 3.4, that is, the “crowding-out effect" is in force. In addition, Proposition 3.7 (2) reveals that where the university makes minimum effort below its potential capacity, a research fund has “nil" effects on teaching effort and student enrollment while it positively affects research effort and research output. More precisely, a research fund does not change any teaching activities and resultant student enrollment of a “small research institute" that has already selected minimum teaching effort. Consolidating all matters discussed, the point is that when a tuition fee is controlled to maximize tuition revenue, research funds may decrease student enrollment, or at best, may be totally ineffective for an increase in student enrollment.

Finally, the discussion facing many government authorities would be whether a change from an existing tuition fee system, where all agents including a university and students have optimized their decisions, is relevant or not. In what follows, we examine the conditions of when research output and student enrollment increase under a controlled tuition fee, accompanied with the change in the tuition fee scheme from a fixed tuition fee.

Proposition 3.8 Suppose that $\hat{e} = (\hat{r}, \hat{t}) = (\frac{k(kF-2s^2)}{k^2-4s^2}, \frac{2s(kF-2s^2)}{k^2-4s^2})$ (with $\hat{r} > 0$, $\hat{t} > 0$, and $\hat{r} + \hat{t} < \overline{a}$), $\hat{n} = \frac{2s(2F-k)}{k^2-4s^2} \in (0, 1)$, and $\hat{R} = [\frac{k(kF-2s^2)}{k^2-4s^2}]^2$, have been initially achieved as a positive interior equilibrium solution under the scheme of a fixed tuition. If the tuition fee scheme has been changed into the scheme of a controlled tuition fee, then:

(1) The maximum effort, $\overline{e} = (\overline{r}, \overline{t})$ with $\overline{r} + \overline{t} = \overline{a}$, $\overline{r} > \hat{r}$, and $\overline{t} > \hat{t}$ can be achieved by: In addition to the condition, $\overline{t} > \hat{t} = \frac{2s(kF-2s^2)}{k^2-4s^2}$,

(1-i) a large capacity $\overline{a}$ that satisfies $\overline{t} > 2\sqrt{k(k-F)}$ and $\frac{k}{2} < F < k$; or

(1-ii) a large research fund that satisfies $F > k$ and $s < \frac{(\sqrt{5}-1)k}{4} \approx 0.309k$;

(2) With regard to student enrollment, $\overline{n} > \hat{n}$ holds for $\overline{t} > \frac{2ks(2F-k)}{k^2-4s^2} = k\hat{n} > \hat{t}$;

(3) With regard to research output, $\overline{R} > \hat{R}$ holds if (1) is the case; and

(4) There can exist particular F and s that induce $\overline{R} > \hat{R}$ and $\overline{n} < \hat{n}$.

We need to take note that when the condition, $\overline{t} > \hat{t} = \frac{2s(kF-2s^2)}{k^2-4s^2}$, is postulated, $\overline{r} > \hat{r}$ is also satisfied from the construction. With this condition in mind, Proposition 3.8 (1) maintains that when a tuition fee is initially fixed, there may be some room to increase both research and teaching effort by applying a flexibly controlled tuition fee that maximizes tuition revenue. Specifically, if the university does not operate at its full capacity under the scheme of a fixed tuition fee, it is possible for us to encourage the university to exert more of its potential capacity. But some additional conditions are necessary for an increase in research and teaching effort, as indicated by Proposition 3.8 (1-i) and (1-ii) which can be immediately led by Proposition 3.6.

First, the potential capacity of the university must be large enough ($\overline{t} > 2\sqrt{k(k-F)}$: “large university") to achieve the maximum effort, $\overline{e} = (\overline{r}, \overline{t})$, when a research fund is small ($F < k$). Otherwise, if the capacity is small ($\bar{t} < 2\sqrt{k(k-F)}$: “small research institute"), the university prefers to choose the minimum effort, $\boldsymbol{e}^0 = (F+\delta_r, \delta_t) \approx (F, 0)$. For the second condition, a large research fund ($F > k$) enables the university to exert maximum research and teaching effort irrespective of its potential capacity. In this instance, the tuition fee level set under the scheme of a fixed tuition fee has to satisfy $s < \frac{k(\sqrt{5}-1)}{4} \approx 0.309k$, which is stricter than $s < \frac{k}{2} = 0.5k$, in order to guarantee that the initial equilibrium solution, $\hat{\boldsymbol{e}} = (\hat{r}, \hat{t})$, is derived from under-enrollment. (Recall that $\hat{n} \in (0, 1)$ is satisfied if F exists in $(\frac{k}{2}, \frac{k}{2} + \frac{k^2-4s^2}{4s})$ as Lemma 3.2 [3] has suggested.)

With regard to Proposition 3.8 (2) to (4), while more research output can be produced if research effort is enhanced along with a change in the tuition fee scheme (Proposition 3.8 [3]), student enrollment cannot be necessarily increased from the initial equilibrium solution, $\hat{n} = \frac{2s(2F-k)}{k^2-4s^2}$ (Proposition 3.8 [2]). The reason for the latter is as follows. Now that the university can freely establish an optimal tuition fee that is adjusted to satisfy $s^* = \frac{t^*}{2}$ under the scheme of a controlled tuition fee, this tuition fee is likely to rise in tandem with improved teaching effort. Although an increase in teaching effort raises the utility of students, a rise in the tuition fee lowers it in an opposite manner. As exhibited in the utility function of students in Equation (3.12), the net effect of an increase in teaching effort on the student utility is $\Delta t^* - \Delta s^* = \frac{\Delta t^*}{2} < \Delta t^*$. This indicates that although improved teaching effort generates higher student enrollment in the scheme of a controlled tuition fee, the degree is smaller than a fixed tuition fee due to an increase in the tuition fee. Hence, a small capacity ($\bar{t} < k\hat{n}$) hinders the university from exceeding the threshold of teaching effort that can achieves a higher student enrollment ($\overline{n} > \hat{n}$). We also need to note that the condition for $\overline{n} > \hat{n}$ is stricter than that for $\bar{t} > \hat{t}$ from the above-mentioned argument.

In view of Proposition 3.8 (4), whether overall student enrollment will increase or decrease in response to the change in the tuition fee scheme depends on parameters, such as university capacity, research fund, initial tuition fee, and mobility cost. Let us focus exclusively on a university capacity and a research fund. If a research fund, F, becomes large, the condition on research output ($\bar{t} > 2\sqrt{k(k-F)}$: decreasing in F) can be more easily satisfied while the condition on student enrollment ($\bar{t} > k\hat{n} = \frac{2ks(2F-k)}{k^2-4s^2}$: increasing in F) is not. This is why the condition on student enrollment may not be maintained in spite of the fact that the condition on research output is satisfied. Subsection 3.8.1 presents a detailed demonstration of Proposition 3.8 (4) by the use of Figure 3.10.

What kind of university would categorically observe an increase in both research output and student enrollment in response to a change made in the tuition fee scheme from “fixed" to “controlled"? As we have already discussed, a “small research institute" that has little capacity may opt for minimum research and teaching effort, $e^0 = (F + \delta_r, \delta_t) \approx (F, 0)$ (Proposition 3.6 [1-ii]). In the case where a research fund, F, is sufficiently large, even a “small research institute" can operate at its full capacity (Proposition 3.6 [2]). However, when the university is still relatively small and the research fund is not sufficient, an increase in student enrollment may not be guaranteed, although research output is likely to increase (Proposition 3.8 [4]). In conclusion, the answer is that only a “large university" with a sufficiently large capacity is expected to enroll more students as well as produce higher research output by a change from a fixed to controlled tuition fee.

3.6. Multiple universities compete for students

This chapter has so far discussed a single university model, but competition is prevalent between universities in reality. In order to continue further extended discussions, Section 3.6 supposes that there are two symmetric universities (denoted by universities 1 and 2, respectively) in the jurisdiction. There is heterogeneity in universities to a varying degree from a realistic viewpoint, such as a payoff (how the university weighs research and teaching) and a capacity. Indeed, we could posit that large and small universities exist in the same jurisdiction. However, the result that we foresee from asymmetric competition in this model might not be very enlightening, because we would end up only observing that a university endowed with a larger capacity or a research fund creates higher research output and student enrollment. By contrast, the symmetric assumption is useful in eliciting by way of a simple analysis what effects competition among universities with an equal footing have on research output and student enrollment. It is also highly likely that more than two universities compete, but examining the case of two universities is enough to obtain some initial basic implications. Even if more than two universities are assumed, it is presumed that the nature of the results would essentially remain the same.

Hereafter, this chapter focuses solely on competition for acquiring students, not research resources. Universities have been recently competing for research grants that governments or private companies offer, but allocating a grant to universities is not always an easy task, especially when the universities are similar in many traits. [28] Since research productivity of the universities is the same depending solely on research effort, it is reasonable to assume that the problem of research fund allocation is beyond the scope of this study despite the possible intention of the financing agency to increase total research output. Accordingly, given a constant amount of a total budget, F, the financing agency simply allocates a research fund in a “pro-rata system": universities 1 and 2 equally gain $\frac{F}{2}$, respectively. [29] In other words, the universities are evenly treated by the financing agency so that they have an equal footing on research activities. The later analysis demonstrates that this pro-rata system could generate an inefficiency of research output.

It is also assumed that both universities 1 and 2 stand at the exact middle point ($\frac{1}{2}$) in the student market, the length of which is 1. It is posited that the universities are not differentiated at all in terms of their locations, but instead, only in their teaching effort levels and tuition fees. Put it simply, the universities compete through these two factors alone. (In this point, when tuition fees are fixed, competition is only through teaching effort.) In general, although universities are dispersed across the country, this analysis intends to consider competition in a more narrow region, for example, at the

[28] In this regard, this analysis looks at competition differently from Gautier and Wauthy (2007), who study research competition among faculty members but not competition for student enrollment.

[29] If there is a large difference in research productivity, the financing agency, which generally intends to maximize research output, will allocate the entire research fund to a more productive university. If the research output function of university i is defined as $R_i = \rho_i r_i b_i$ where ρ_i is a parameter of research efficiency (research productivity can be represented as $\rho_i r_i$), the financing agency allocates a research fund only to a university that has a larger ρ_i. In the case where R_i is a concave function with regard to b_i as supposed by Gautier and Wauthy (2007), research funds are likely to be allocated to multiple universities with consideration given to a more productive university.

level of cities and states. [30] To be exact, we need to keep in mind that this assumption is different from the reality developed in the empirical section (Section 3.2).

Lastly, we suppose that if the teaching effort levels and tuition fees are equivalent between the two universities, they evenly divide a mass of students who have applied for enrollment. For descriptive purposes, we may as well think about a simplified situation where considering all factors being equal, universities 1 and 2 gain the left and right half of the student market, respectively.

3.6.1. Competition under fixed tuition fees

Consider first that the tuition fees are fixed by regulation. Because students are permitted to enroll at only one university and so will select the one that can provide them with the highest possible utilities, universities 1 and 2 produce the same teaching effort under the terms of competitiveness. Otherwise, one university, which exerts less teaching effort than the other, is likely to lose all the students they have admitted so far, and as a result, is faced with negative payoffs. More generally, a particular student having a utility, $u = t - s - k\left(\frac{1}{2} - x\right)$ located at $x < \frac{1}{2}$, will choose a university that offers higher teaching standards given a constant tuition fee and mobility cost. (When $x > \frac{1}{2}$ holds, the same argument is applied, too.) This argument leads to the result that teaching effort of the two universities is likely to be equal at an equilibrium: $t = t_1 = t_2$. Accordingly, the universities also accept the same number of students evenly dividing the student market: $n_1 = n_2 = \frac{s(t-s)}{k}$.

[30] It seems that Boston and Cambridge in the U.S., which is a famous jurisdiction where universities agglomerate (for example, Harvard University, MIT, Boston University, among others), exemplifies competition between universities at the city and/or state level.

What we need to show is that when the first-order conditions with regard to r and t are satisfied, both universities gain strictly positive payoffs: $U_1 > 0$ and $U_2 > 0$. Focusing on university 1, we consider the following maximization problem:

$$\max_{r_1,t_1} U_1(r_1,t_1) = r_1 b_1 - \left(\frac{r_1^2}{2} + \frac{t_1^2}{2}\right) \text{ subject to } b_1 = \frac{F}{2} + \frac{s(t_1 - s)}{k}. \tag{3.37}$$

From Equation (3.37), the first-order conditions are provided by $r_1 = \frac{F}{2} + \frac{s(t_1-s)}{k}$ and $t_1 = (\frac{s}{k})r_1$, respectively. Suppose that under-enrollment in the entire student market is still prevalent (i.e. $n = n_1 + n_1 \in (0,1)$) and that the capacity of each university is not binding at positive research and teaching effort (i.e. $r_i > 0$, $t_i > 0$, and $r_i + t_i < a$ for $i = 1,2$). Then, research and teaching effort, student enrollment, and research output of university 1 are derived as follows: $(\hat{r}_1, \hat{t}_1) = (\frac{k(kF-2s^2)}{2(k^2-s^2)}, \frac{s(kF-2s^2)}{2(k^2-s^2)})$, $\hat{n}_1 = \frac{s(F-2k)}{2(k^2-s^2)}$, and $\hat{R}_1 = [\frac{k(kF-2s^2)}{2(k^2-s^2)}]^2$, where $F > 2k > \frac{2s^2}{k}$ is assumed for $\hat{r}_1 > 0$, $\hat{t}_1 > 0$, $\hat{n}_1 > 0$, and $\hat{R}_1 > 0$. The second-order condition, $k^2 - s^2 > 0 \Leftrightarrow s < k$, is also required, but this condition is always satisfied under $s < \frac{k}{2}$, which was assumed in the case of a single university.

In addition, university 1 obtains a payoff, $\hat{U}_1 = \hat{R}_1 - (\frac{\hat{r}_1^2}{2} + \frac{\hat{t}_1^2}{2}) = \frac{(kF-2s^2)^2}{8[k^2-s^2]} > 0$, which is strictly positive under these assumptions. We can also derive $\hat{U}_2 > 0$ for university 2 from the symmetry. Hence, when universities 1 and 2 compete for student enrollment through teaching effort, it is expected that they have to further increase their teaching effort, t_1 and t_2, up to the point where they cannot earn any positive payoffs from their activities; that is, they determine their teaching effort so as to satisfy the zero-payoff condition, $\hat{U}_1 = \hat{U}_2 = 0$. [31]

[31] The reasoning employed here is analogous to the mechanism working in Bertrand competition.

It is possible that the universities may be inclined to decrease research effort to divert it into teaching activities, aiming at winning competition in the student market. If we take into consideration such behavior of the universities, multiple equilibria can occur with regard to research and teaching effort. In a multiple equilibria case, the analysis would be generally more complicated, so that we find it difficult to specify a particular set of an equilibrium solution regarding research and teaching effort. In practice, many combinations of research and teaching effort made by universities 1 and 2 can be an equilibrium solution. (Symmetry may not be always guaranteed either.) In order to determine uniquely an equilibrium, Assumption 3.2 is still made that research effort is always adjusted to satisfy the first-order condition, $r_i = \frac{F}{2} + \frac{s(t_i - s)}{k}$ $(i = 1, 2)$, for any given t_i.

Based on these settings, let us search for an equilibrium solution that generates zero-payoffs to the universities. In other words, we aim to identify $e_i^{C1} = (r_i^{C1}, t_i^{C1})$ that satisfies $U_i(r_i^{C1}, t_i^{C1}) = r_i^{C1}[\frac{F}{2} + \frac{s(t_i^{C1} - s)}{k}] - [\frac{(r_i^{C1})^2}{2} + \frac{(t_i^{C1})^2}{2}] = 0$ for $i = 1, 2$. The following proposition indicates the equilibrium solution and compares with those obtained in the case of a single university.

Proposition 3.9 Suppose that two universities 1 and 2 are allocated a research fund, $\frac{F}{2}$, respectively, and compete for student enrollment through their teaching effort under the scheme of a fixed tuition fee. Assuming that (a) under-enrollment occurs ($n_1^{C1} + n_2^{C1} \in (0, 1)$), (b) the capacity constraint is not binding at a positive interior equilibrium solution ($r_i^{C1} > 0$, $t_i^{C1} > 0$, and $r_i^{C1} + t_i^{C1} < a_i$ for $i = 1, 2$), and (c) teaching effort cannot be increased at the expense of research effort (Assumption 3.2: $r_i^{C1} = \frac{F}{2} + \frac{s(t_i^{C1} - s)}{k}$ for

$i = 1, 2$ holds for any t_i^{C1}), we obtain the following results:

(1) $r_1^{C1} = r_2^{C1} = \frac{kF-2s^2}{2(k-s)}$; and $r^{C1} = r_1^{C1} + r_2^{C1} = \frac{kF-2s^2}{k-s} > \hat{r}$ for $s < \frac{k}{4} = 0.25k$;

(2) $t_1^{C1} = t_2^{C1} = \frac{kF-2s^2}{2(k-s)}$; and $t^{C1} = t_1^{C1} + t_2^{C1} = \frac{kF-2s^2}{k-s} > \hat{t}$ for $s < \frac{(\sqrt{3}-1)k}{2} \approx 0.366k$;

(3) $n_1^{C1} = n_2^{C1} = \frac{F-2s}{2(k-s)}$; and $n^{C1} = n_1^{C1} + n_2^{C2} = \frac{F-2s}{k-s} > \hat{n}$ for $s < \frac{k}{4} = 0.25k$; and

(4) $R_1^{C1} = R_2^{C1} = [\frac{kF-2s^2}{2(k-s)}]^2$; and $R^{C1} = R_1^{C1} + R_2^{C1} = \frac{1}{2}(\frac{kF-2s^2}{k-s})^2 < \hat{R}$ for every F, s and, k.

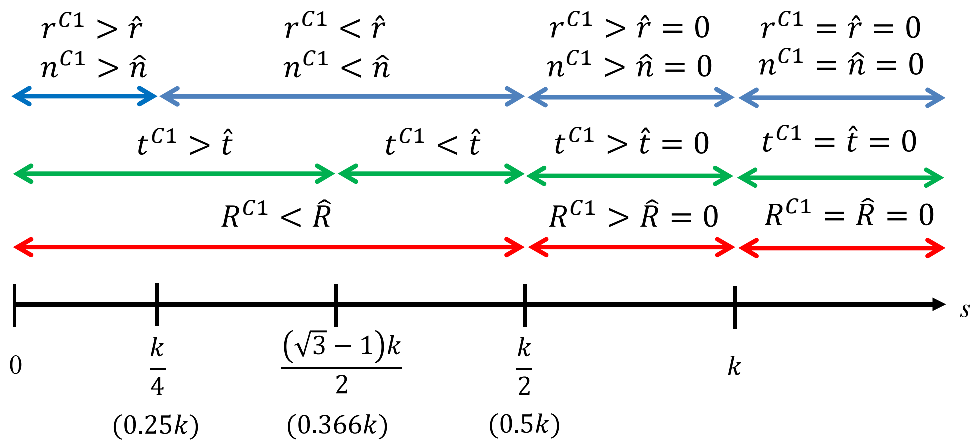


Figure 3.8. Comparison of equilibrium variables.

Figure 3.8 illustrates the summary of Proposition 3.9 along with the relationship between s and k. What we should note is that competition between universities 1 and 2 does not necessarily generate higher student enrollment collectively than that of a single university, even when the sum of the teaching effort by the universities are increased more than that of a single university. That is, $t^{C1} > \hat{t}$ and $n^{C1} < \hat{n}$ hold for $s \in (\frac{k}{4}, \frac{(\sqrt{3}-1)k}{2})$, which is pertinent to somewhat seemingly incoherent situation. This is largely attributable to adjacency of university location that causes exceedingly fierce competition between the universities. More precisely, a significant amount of teaching effort has a positive effect of stimulating student enrollment, but a certain amount is wasted in a sense that the universities, which are located at the adjacent point, conduct overlapping investment in teaching activities.

In particular, Proposition 3.9 (2) states that when a tuition fee is set relatively high, a single university can anticipate a more positive effect among the tuition fee, teaching effort, and student enrollment (i.e. $\frac{\partial \hat{t}}{\partial s} > 0$ and $\frac{\partial \hat{n}}{\partial s} > 0$ from Proposition 3.4 [2]), and therein, gaining a large benefit from monopoly of the student market. Hence, in the range of a relatively high tuition fee ($s \in (\frac{(\sqrt{3}-1)k}{2}, \frac{k}{2})$), a single university is expected to produce higher student enrollment while multiple universities still suffer from wasteful duplication of teaching effort. By contrast, let us suppose that a tuition fee is set relatively low ($s \in (0, \frac{k}{4})$). Whereas the positive tuition fee effect accruing to a single university is not conspicuous, the positive effect of teaching competition becomes relatively strong despite duplicative teaching effort, and therefore, multiple universities create higher total student enrollment than a single university.

Furthermore, notably enough, the sum of research output cannot overwhelm that produced by a single university for any level of research funds, tuition fees, and mobility costs, in spite of the result that the sum of research effort is increased in the case of $s \in (0, \frac{k}{4})$ (Proposition 3.9 [1] and [4]). The key to solving the question lies in the fact that research output explicitly exhibits increasing returns to scale since the research output function can be generally denoted by $R = rb = r^2$. [32] Consequently, the two universities cannot exploit economies of scale to the extent of a single university since each research effort made by them is generally smaller than that of a single

[32] When the degree of substitutability is zero ($\varepsilon = 0$), $r = b$ holds by the first-order condition in Equation (3.15). Since research output function is defined by $R = rb$, we finally obtain $R = r^2$.

university. [33] It follows that it is more difficult for multiple universities to produce research output in total than a single university due to the lower level of economies of scale, which becomes much more conspicuous as the number of universities gets large given a total research fund evenly provided by the financing agency.

Indeed, although this result heavily relies on the characteristics of increasing returns to scale, it is quite evident in the science fields that research activities typically generate “positive feedbacks". For example, as investment in research facilities raise the chance of making achievements, which in turn give researchers more chances to improve research facilities, research output increases with a positive feedback loop (David, 1993). (Subsection 3.8.3 investigates whether the research output function exhibits increasing returns to scale in accordance with the shape of the cost function.)

To recap, introducing another university into the model on the assumption that a tuition fee is fixed and a given research fund is evenly divided, does not always induce a favorable result in terms of both research output and student enrollment. While student enrollment will be increased when the tuition fee is initially set at a lower level, it is difficult to hope for more research output under the assumed situation.

3.6.2. Competition under controlled tuition fees

Under the scheme of controlled tuition fees, universities 1 and 2 have alternatives of lowering their tuition fees as well as increasing teaching effort in order to obtain higher student enrollment in student market competition. Provided that tuition fees are

[33] $\hat{r} - r_i^{C1} = \frac{k(kF-2s^2)}{k^2-4s^2} - \frac{kF-2s^2}{2(k-s)} = \frac{(kF-2s^2)[(k-s)^2+2s^2]}{2(k^2-4s^2)(k-s)} > 0 \Leftrightarrow \hat{r} > r_i^{C1}$ for $i = 1, 2$.

reduced, the universities do not always maximize tuition revenue, E, any longer, and whence, there cannot be a unique equilibrium solution. As examined in Proposition 3.6, we need to recall that when a single university operates in the scheme of a controlled tuition fee and is given an amount of a research fund, $F \in (\frac{k}{2}, k)$, there are two possible equilibrium solutions of research and teaching effort, that is, $\overline{e} = (F + \frac{\bar{t}^2}{2k}, \bar{t})$ with $\bar{r} + \bar{t} = \bar{a}$ (large university) and $e^0 = (F + \delta_r, \delta_t) \approx (F, 0)$ (small research institute), in accordance with the capacity of the university.

Now assume that there are two symmetric “large universities" (universities 1 and 2) with an evenly distributed research fund. Before the competition starts, the two universities use their full capacities. What happens in this sort of an environment? As mentioned at the beginning of this section, since one university (denoted by university 1) will begin to commit itself to intensifying its teaching effort at the expense of research effort or lowering its tuition fee in order to attract more students, the other university (denoted by university 2) can no longer maintain the initial choice. In turn, university 2 will be also compelled to increase its teaching effort or to lower its tuition fee in an analogous manner with university 1. Competition will continue until the payoffs of the two universities converge to a minimum level (as in the previous case, generally zero) and their effort will reach the capacity constraints ($r_i^* + t_i^* = \bar{a}$). [34] Such competition through both an increase in teaching effort and a reduction in a tuition fee would generate higher student enrollment than a single university. However, a much more complicated problem is that universities 1 and 2 simultaneously determine both

[34] The current explanation of the process in competition between the universities is based on a dynamic game context, but it is just for a descriptive purpose. An equilibrium can be achieved in a one-shot static game, where the universities immediately anticipate it at the beginning of the game.

effort levels and tuition fees. Accordingly, since there might be practically vast combinations of an equilibrium where the payoffs of the universities converge to zero and effort reaches a capacity constraint, it is difficult to uniquely identify the levels of research and teaching effort as well as tuition fees.

More interesting thing is to analyze the case of two “small research institutes" that receive an equivalent amount of a research fund. If a single small research institute exists in the jurisdiction, it is likely to operate at the lowest teaching level: $\boldsymbol{e}^0 \approx (F, 0)$, $s^0 \approx 0$, $n^0 \approx 0$, and $R^0 \approx F^2$, as demonstrated in Proposition 3.6. In this environment, provided that there are two small research institutes in the same jurisdiction, the conjecture is that the sum of student enrollment is sure to increase from approximately zero to a significantly positive value due to teaching competition in the student market. And yet, it is ambiguous whether the sum of research output increases or decreases, not only due to the characteristics of scale economies, but also due to a decline in research effort.

In order to derive one explicit combination of the equilibrium solution, suppose again Assumption 3.2 that the first-order condition of research effort is always satisfied given any teaching effort likewise in the previous subsection. On top of that, a restriction is placed on the tuition fee setting by making the following assumption.

Assumption 3.3 The government authority forces universities 1 and 2 to set their tuition fees that maximize tuition revenue of each university.

This assumption leads to the competition only through teaching effort, but not the tuition fee. In conjunction with Assumption 3.2, university i for $i = 1, 2$ determines the direction of its research and teaching effort, $\boldsymbol{e}_i = (r_i, t_i)$ that satisfies $r_i(t_i) = \frac{F}{2} + \frac{t_i^2}{2k}$. It is also expected that the universities will eventually choose research and teaching effort, $\overline{e}' = (\overline{r}', \overline{t}')$ and $\overline{r}' + \overline{t}' = \overline{a}$, where they exploit their capacities fully due to teaching competition. [35] To come right to the point, this so-called "partially regulated competition" solely via teaching effort may be more desirable from the viewpoint of both research output and student enrollment than a single university that opts for the minimum level of activities.

Proposition 3.10 Suppose that (a) tuition fees are regulated so as to maximize tuition revenue (Assumption 3.3: $s_i^{C2} = \frac{t_i^{C2}}{2}$ for $i = 1, 2$) and (b) teaching effort cannot be increased at the expense of research effort (Assumption 3.2: $r_i^{C2} = \frac{F}{2} + \frac{s_i^{C2}(t_i^{C2} - s_i^{C2})}{k}$ for $i = 1, 2$). Universities 1 and 2 are assumed to be granted with an evenly divided research fund ($\frac{F}{2}$ for each), to have small capacities ($\overline{t}' < 2\sqrt{k(4k - F)}$), and to choose $\tilde{\boldsymbol{e}}^0 = (\frac{F}{2} + \delta_r, \delta_t) \approx (\frac{F}{2}, 0)$ without the presence of teaching competition. If universities 1 and 2 compete for student enrollment only through teaching effort, then:

(1) The equilibrium solution achieves:

(1-i) When $U_i(\overline{r}', \overline{t}') \geq 0$ holds, $\boldsymbol{e}_i^{C2} = \overline{\boldsymbol{e}}' = (\overline{r}', \overline{t}') = (\frac{F}{2} + \frac{(\overline{t}')^2}{4k}, \overline{t}')$, $s_i^{C2} = \frac{\overline{t}'}{2}$, $n_i^{C2} = \frac{\overline{t}'}{2k}$, and $R_i^{C2} = \frac{1}{4}(F + \frac{(\overline{t}')^2}{2k})^2$ for $i = 1, 2$, where $(\overline{r}', \overline{t}')$ satisfies $\overline{r}' + \overline{t}' = \overline{a}$, $\overline{r}' = \frac{F}{2} + \frac{(\overline{t}')^2}{4k}$, and $\overline{t}' < 2\sqrt{k(4k - F)}$;

(1-ii) When $U_i(\overline{r}', \overline{t}') < 0$ holds, $\boldsymbol{e}_i^{C2} = \boldsymbol{e}' = (r', t') = (\frac{F}{2} + \frac{(t')^2}{4k}, t')$, $s_i^{C2} = \frac{t'}{2}$, $n_i^{C2} = \frac{t'}{2k}$, and $R_i^{C2} = \frac{1}{4}[F + \frac{(t')^2}{2k}]^2$ for $i = 1, 2$, where (r', t') satisfies $\frac{F}{2} < r' < \overline{r}'$, $0 < t' < \overline{t}' <$

---

[35] Notice that $\overline{e}' \neq \overline{\boldsymbol{e}}$ generally holds because the current reaction function, $r_i(t_i) = \frac{F}{2} + \frac{t_i^2}{2k}$, is different from $r(t) = F + \frac{t^2}{2k}$.

$2\sqrt{k(4k-F)}$, $r' = \frac{F}{2} + \frac{(t')^2}{4k}$, and $U_i(r', t') = 0$;

(2) Comparing the equilibrium solutions between the cases of multiple small research institutes and a single small research institute that opts for $\boldsymbol{e}^0 = (F + \delta_r, \delta_t) \approx (F, 0)$, $s^0 \approx 0$, $n^0 \approx 0$, and $R^0 \approx F^2$ under $\bar{t} < 2\sqrt{k(k-F)}$ and $F < k$, we obtain:

(2-i) $r^{C2} = r_1^{C2} + r_2^{C2} > r^0$, $t^{C2} = t_1^{C2} + t_2^{C2} > t^0$, and $n^{C2} = n_1^{C2} + n_2^{C2} > n^0$ for both cases of (1-i) and (1-ii); and

(2-ii) $R^{C2} = R_1^{C2} + R_2^{C2} > R^0$ for the case of (1-i) if $\bar{t}'$ is sufficiently close to $2\sqrt{k(4k-F)}$ (i.e. $\bar{t}' \to 2\sqrt{k(4k-F)}$).

Figure 3.9 depicts the equilibrium solutions for Proposition 3.10 (1-i) and (1-ii). If the payoff obtained from the full capacity is more than zero ($U_i(\bar{r}', \bar{t}') \geq 0$), competition through teaching effort between the two universities induces the equilibrium solution to move further to the upper right toward the point of the maximum effort along the reaction curve (Point B of Figure 3.9 [i]). On the other hand, if this payoff is negative ($U_i(\bar{r}', \bar{t}') < 0$), such competition will end before the capacity is attained, where the payoff reaches exactly zero (Point C of Figure 3.9 [ii]). Since the utility at the minimum effort level is positive ($U_i(\frac{F}{2}, 0) = \frac{F^2}{8} > 0$), we can necessarily find a set of research and teaching effort, $\boldsymbol{e}' = (r', t')$, where $\frac{F}{2} < r' < \bar{r}'$ and $0 < t' < \bar{t}' < 2\sqrt{k(4k-F)}$, which generates a zero payoff, $U_i(r', t') = 0$ (the intermediate-value theorem).

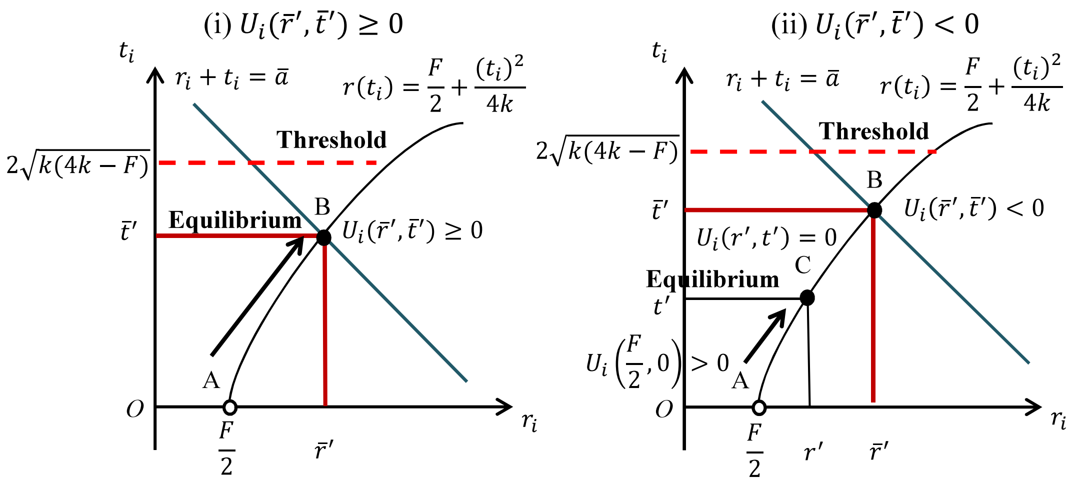


Figure 3.9: Equilibrium in teaching competition between small research institutes.

Proposition 3.10 (2-i) reveals that even if a single university does not exert its full capacity, competition in the student market is sure to create higher student enrollment. Most notably, Proposition 3.10 (2-ii) also demonstrates that when the university capacities are in close proximity to the threshold that determines the equilibrium solution of universities 1 and 2 in the absence of teaching competition, research output is also expected to increase in comparison with a single university. This suggests that the positive impact of increasing research and teaching effort overwhelms the negative impact of losing an increasing return in research output.

The important assumption behind these results is that the government prevents the universities through regulatory policies from “dumping" their tuition fees. As such, the equilibrium solution which we have focused on seems picked arbitrarily. But the education system is also arbitrarily regulated in reality. This may be supported by the fact that tuition fee guidelines are introduced in some countries, which can be interpreted as a way to maintain research activities by preventing the university budget from being undermined, as well as a way to protect students from a sharp rise in tuition fees in the absence of teaching improvements. [36] Based on such a background, our model selects a particular equilibrium solution via introducing additional constraints.

3.7. Concluding remarks

Chapter 3 examined how mutually-connected research and teaching activities of universities interact to generate research output and student enrollment, based on a setting where universities obtain a research fund from a governmental financing agency and earn tuition revenue from students by setting their tuition fee.

In the first place, this chapter derived empirical implications on the basis of U.S. university data. It aimed to primarily investigate what effect on research output and student enrollment works in response to an increase in a research fund. Although the result must be interpreted in a careful manner, it was demonstrated that a research fund allocated to private universities is ineffective or might be negatively related to the number of doctorates awarded possibly due to strong substitutability of research and teaching activities. Additionally, this chapter also found that a research fund might have a negative relation with student enrollment of private universities, which implies that substitutability or the crowding-out effect resulting from small capacities may also be in force in the case of private universities. Hence, we need to remind ourselves that policymakers and university administrative officials should undertake

[36] For example, in 2002 the Japanese government prescribed a desirable tuition fee level as a standard with which national universities need to conform, and most national universities set their tuition fees taking into account the government's intended purpose.

efforts to mitigate substitutability and eliminate their capacity constraints that could cause a crowding-out effect.

Based upon the empirical observations, we theoretically argued that substitutability between research and teaching activities is of great importance especially when considering the way a research fund affects research output and student enrollment, which is a main focus of this chapter. More precisely, it was demonstrated that if the degree of substitutability is large enough, not only student enrollment but also research output could decrease due to an increase in a research fund. The intuition of this is as follows; since strong substitutability may drastically decrease teaching effort and student enrollment in response to an increase in research effort caused by an incremental research fund, a smaller research budget may result in decreased research output in the end. This finding that an increased research fund could have a negative effect on research output seems much paradoxical. However, policymakers and university administrative officials should keep this possibility in mind when they intend to have universities yield higher research output by enhancing research funds.

Assuming that the degree of substitutability is zero in an illustrative model case, we found that the results significantly vary according to whether the tuition fee is fixed or controlled. In the case of a fixed tuition fee, whereas research funds can increase both research output and student enrollment when the capacity of the university to engage in research and teaching activities is not fully utilized (“multiplier effect"), it crowds out student enrollment when the university operates at its full capacity (“crowding-out effect"). This simple result derives from the nature of a university that is evaluated ultimately by research output, not teaching outcome. Consequently, if the university does not have an extra margin of its capacity to work with, every additional activity can be devoted into research activities.

By contrast, in the case where the tuition fee is controlled by a government authority in maximizing tuition revenue, a marginal amount of a research fund never has any positive effect on student enrollment due to the emergence of a “binary divide" of the university (namely, multiple equilibria). More precisely, this “binary divide" implies that while a university with a large capacity (“large university") operates at its full capacity, a university with a small capacity (“small research institute" or “small college") opts for marginal activities. In these two cases, an increased research fund leads only to an increase in research output, but not student enrollment, and with regard to a “large university", the crowding-out effect operates to decrease student enrollment. With this in mind, this chapter revealed that in order to make a “small research institute" grow from engaging in marginal activities, providing a sufficiently large amount of a research fund or enhancing the capacity of the university is required.

As an extensive discussion, this chapter also analyzed the effect of competition among two universities that are seeking an increase in student enrollment, and compares the results between single and multiple universities. The analysis suggested that in both schemes of a tuition fee, partially regulated competition only through teaching effort can increase overall student enrollment given the same total amount of research funds evenly divided to the universities. But it was also proven that the sum of research output may be decreased by competition because the effect of increasing returns to scale in research output is lost.

In conclusion, the answer to the question, “Can we simultaneously increase research output and student enrollment?", which is an issue referred to in the title of this paper, could be “No" depending on certain conditions. In one case, student enrollment may be decreased while research output is increased by the existence of research funds. In the other extreme case, when the degree of substitutability between research and teaching activities is strong enough, even research output may be decreased in response to an increase in research funds.

The issues to be further scrutinized are briefly described in what follows. First, although the financing agency in this model only allocates constant research funds to universities, we can consider a dynamic model where it depends on research productivity or research output, so that the decision of the financing agency is also endognized. Second, in relation to the above, the multiple universities model can be modified to include competition for research funds as well as students. It is expected from this modification that heterogeneous universities are divided into research and teaching specific universities pointed out by Del Rey (2001) and De Fraja and Iossa (2002). This change in the model setting is likely to consequentially affect total research output and student enrollment. Lastly, in order to make the empirical model closer to reality, it is necessary to introduce measures such as the number of articles and patents for appropriate indicators of research output, research and teaching quality, and other significant explanatory variables.

3.8. Appendices

3.8.1. Proofs of Propositions and Lemmas

The mathematical demonstrations are gathered in this subsection. The proof of Propositions and Lemmas are as follows.

Proposition 3.1 (1) $\frac{\partial r^*}{\partial F} = \frac{1}{|A_F|}[\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})] > 0$ because $\frac{\partial^2 C}{\partial t^2} > 0$ and $\frac{\partial^2 n}{\partial t^2} \leq 0$ from the assumptions.

(2-i) When $\frac{\partial t^*}{\partial F} = \frac{1}{|A_F|}[s(\frac{\partial n}{\partial t}) - \frac{\partial^2 C}{\partial r \partial t}] > 0$ holds, it is obvious that $\frac{\partial^2 C}{\partial r \partial t} < s(\frac{\partial n}{\partial t})$. As for student enrollment, $\frac{\partial n^*}{\partial F} = (\frac{\partial n}{\partial t})(\frac{\partial t^*}{\partial F}) > 0$ because $\frac{\partial t^*}{\partial F} > 0$. Since $R^* = r^*(F + sn^*)$, we can derive $\frac{\partial R^*}{\partial F} = (\frac{\partial r^*}{\partial F})(F + sn^*) + r^*[1 + s(\frac{\partial n^*}{\partial F})] > 0$ in the above condition.

(2-ii) By the same derivation of (2-i), if $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t})$ holds, $\frac{\partial t^*}{\partial F} < 0$ and $\frac{\partial n^*}{\partial F} < 0$.

(3) Transforming the condition for $\frac{\partial R^*}{\partial F} < 0$, we obtain:

$$r^* + \frac{\partial r^*}{\partial F}(F + sn^*) + r^* s \left[\frac{\partial n}{\partial t}\frac{1}{|A_F|}\left[s\left(\frac{\partial n}{\partial t}\right) - \frac{\partial^2 C}{\partial r \partial t}\right]\right] < 0$$

$$\iff \frac{\partial^2 C}{\partial r \partial t} > s\left(\frac{\partial n}{\partial t}\right) + \underbrace{\frac{|A_F| r^* + \left[\frac{\partial^2 C}{\partial t^2} - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right](F + sn^*)}{r^* s(\frac{\partial n}{\partial t})}}_{J} = s\left(\frac{\partial n}{\partial t}\right) + J, \tag{3.38}$$

where $J = \frac{|A_F| r^* + \left[\frac{\partial^2 C}{\partial t^2} - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right](F + sn^*)}{r^* s(\frac{\partial n}{\partial t})}$. By assuming the case of $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t}) > 0$, we find that the determinant is positive: $|A_F| > 0$. It should be noted that because $|A_F|$ in Equation (3.38) also includes $\frac{\partial^2 C}{\partial r \partial t}$, we need to check whether this inequality still holds. While the left-hand side of Equation (3.38) is increasing in $\frac{\partial^2 C}{\partial r \partial t}$, the right-hand side is decreasing in $\frac{\partial^2 C}{\partial r \partial t}$ because $|A_F| = (\frac{\partial^2 C}{\partial r^2})[\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})] - [\frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t})]^2$ is decreasing in $\frac{\partial^2 C}{\partial r \partial t}$ for $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t})$. In addition, provided that $\frac{\partial^2 C}{\partial r \partial t} = s(\frac{\partial n}{\partial t})$, the right-hand side of Equation (3.38) is equivalent to:

$$s\left(\frac{\partial n}{\partial t}\right) + \underbrace{\frac{r^*[\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t})] + [\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})](F + sn^*)}{r^* s(\frac{\partial n}{\partial t})}}_{\text{positive}} > s\left(\frac{\partial n}{\partial t}\right) \tag{3.39}$$

from the assumptions. Hence, we can find a particular point, $\frac{\partial^2 C}{\partial r \partial t} = s(\frac{\partial n}{\partial t}) + \Omega$ with $\Omega > 0$, which leads to $\frac{\partial^2 C}{\partial r \partial t} = s(\frac{\partial n}{\partial t}) + \frac{|A_F| r^* + [\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})](F + sn^*)}{r^* s(\frac{\partial n}{\partial t})}$. Accordingly, Ω exactly corresponds to $\Omega = J = \frac{|A_F| r^* + [\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})](F + sn^*)}{r^* s(\frac{\partial n}{\partial t})} > 0$ in Equation (3.38). In conclusion, $\frac{\partial R^*}{\partial F} < 0$ for $\frac{\partial^2 C}{\partial r \partial t} > s(\frac{\partial n}{\partial t}) + \Omega$ is established. ■

Lemma 3.1 Equation (3.16) can be transformed into $t = (\frac{2s}{k} - \varepsilon)r$. By substituting this t into in Equation (3.15), we obtain $r = (\frac{2s}{k} - \varepsilon)^2 r + \frac{kF - 2s^2}{k}$. Solving this equation with regard to r provides $\hat{r} = \frac{k(kF - 2s^2)}{(1 - \varepsilon^2)k^2 + 4ks\varepsilon - 4s^2}$. Also, $\hat{t} = \frac{(2s - k\varepsilon)(kF - 2s^2)}{(1 - \varepsilon^2)k^2 + 4ks\varepsilon - 4s^2}$ can be derived from these two equations. It can be shown that $\hat{r}$ and $\hat{t}$ are strictly positive under the following assumptions, $(1 - \varepsilon^2)k^2 + 4ks\varepsilon - 4s^2 > 0$, $kF - 2s^2 > 0$, and $2s - k\varepsilon > 0$. These conditions are summarized into $F > \frac{2s^2}{k}$ and $-1 + \frac{2s}{k} < \varepsilon < \frac{2s}{k}$. ■

Proposition 3.2 $\frac{\partial r^*}{\partial \varepsilon} = \frac{2k^2(kF - 2s^2)(k\varepsilon - 2s)}{[(1 - \varepsilon^2)k^2 + 4ks\varepsilon - 4s^2]^2} < 0$ because $F > \frac{2s^2}{k}$ and $\varepsilon \in (-1 + \frac{2s}{k}, \frac{2s}{k})$ are assumed in an interior equilibrium solution. Since Equation (3.16) implies $t^* = (\frac{2s}{k} - \varepsilon)r^*$, we can derive $\frac{\partial t^*}{\partial \varepsilon} = -r^* + (\frac{2s}{k} - \varepsilon)\frac{\partial r^*}{\partial \varepsilon} < 0$ because $\frac{2s}{k} - \varepsilon > 0$ and $\frac{\partial r^*}{\partial \varepsilon} < 0$. Student enrollment is represented as $n^* = \frac{2(t^* - s)}{k}$, and hence, $\frac{\partial n^*}{\partial \varepsilon} = (\frac{2}{k})\frac{\partial t^*}{\partial \varepsilon} < 0$. Finally, noting that $R = r^*(F + sn^*)$, we obtain $\frac{\partial R^*}{\partial \varepsilon} = (\frac{\partial r}{\partial \varepsilon})(F + sn^*) + r^* s(\frac{\partial n^*}{\partial \varepsilon}) < 0$. ■

Lemma 3.2 Based on Equation (3.20), $\hat{n} = \frac{2(\hat{t} - s)}{k} = \frac{2}{k}[\frac{2s(kF - 2s^2)}{k^2 - 4s^2} - s] = \frac{2s(2F - k)}{k^2 - 4s^2}$. Since Equation (3.15) indicates $\hat{r} = b$, we derive $\hat{R} = \hat{r}b = \hat{r}^2 = [\frac{k(kF - 2s^2)}{k^2 - 4s^2}]^2$.

The condition of student enrollment requires $0 \leq \hat{n} \leq 1 \Leftrightarrow 0 \leq \frac{2(t-s)}{k} \leq 1 \Leftrightarrow 0 \leq \frac{2s(2F-k)}{k^2-4s^2} \leq 1$. Solving these inequalities for t and F, we obtain $s \leq t \leq s + \frac{k}{2}$ and $\frac{k}{2} \leq F \leq \frac{k}{2} + \frac{k^2-4s^2}{4s}$, respectively. Clearly, these two conditions coincide with each other. W will see hereafter the condition regarding F.

In the first place, let us consider the case, $\frac{2s^2}{k} < F \leq \frac{k}{2}$, where the university needs to decide whether to undertake significantly positive research and teaching effort or not. More precisely, the university chooses either minimum teaching effort that assures an infinitesimal number of student enrollment (i.e. $\boldsymbol{e}^* = (F + \delta_r, s + \delta_t) \approx (F, s)$) or nil (i.e. $\boldsymbol{e}^* = (0, 0)$), taking into account the payoffs obtained from them. By approximate calculation, the payoff becomes $U(F, s) = \frac{F^2-s^2}{2}$ and $U(0, 0) = 0$, respectively. It can be demonstrated that $U(F, s) < U(0, 0)$ if and only if $F < s$. Hence, if $\frac{2s^2}{k} < F < s$ holds, the equilibrium solution is $\boldsymbol{e}^* = (0, 0)$, $n^* = 0$, and $R^* = 0$ (statement [1]). Otherwise, if $s < F \leq \frac{k}{2}$, we obtain $\boldsymbol{e}^* = (F + \delta_r, s + \delta_t) \approx (F, s)$, $n^* = \delta_n \approx 0$, and $R^* = (F + \delta_r)^2 \approx F^2$ considering Assumption 3.1 (statement [2]). Next, consider $F \geq \frac{k}{2} + \frac{k^2-4s^2}{4s}$, where the university enrolls all students in the jurisdiction ($n^* = 1$). In this case, the teaching effort, $t^* = s + \frac{k}{2}$, is chosen at the right corner, and thereby, the university budget amounts to $b = s + F$ (tuition revenue is $n^* s = s$ for $n^* = 1$). Then, the university gains $U(s + F, s + \frac{k}{2}) = \frac{[2(s+F)]^2-(2s+k)^2}{8}$. Comparing the utility at $\boldsymbol{e} = (0, 0)$, we can derive $U(s + F, s + \frac{k}{2}) > U(0, 0) = 0$ for $F > \frac{k}{2}$. But $F > \frac{k}{2}$ is always satisfied for the setting, $F \geq \frac{k}{2} + \frac{k^2-4s^2}{4s}$. Because we can conclude $U(s+F, s+\frac{k}{2}) > U(0, 0)$, $\mathbf{e}^* = (s+F, s+\frac{k}{2})$, $n^* = 1$, and $R^* = (s+F)^2$ for $F \geq \frac{k}{2} + \frac{k^2-4s^2}{4s}$ is an equilibrium solution (statement [4]). Finally, when $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2-4s^2}{4s}$, the university obtains the payoff, $U(\hat{r},\hat{t}) = \hat{R} - (\frac{\hat{r}^2}{2} + \frac{\hat{t}^2}{2}) = \frac{1}{2}[\frac{(kF-2s^2)^2}{k^2-4s^2}] > 0$. Hence, the equilibrium solution is $e^* = (\hat{r},\hat{t})$, $n^* = \hat{n} \in (0,1)$, and $R^* = \hat{R}$ for $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2-4s^2}{4s}$ (statement [3]). ■

Proposition 3.3 (1) When $\frac{2s^2}{k} < F < s$ holds, F does not affect research effort and research output, so that $\frac{\partial r^*}{\partial F} = 0$ and $\frac{\partial R^*}{\partial F} = 0$. On the other hand, when $F > s$ holds, we find r^* and R^* increasing in F, and hence, obtain $\frac{\partial r^*}{\partial F} > 0$ and $\frac{\partial R^*}{\partial F} > 0$.

(2) When $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2-4s^2}{4s}$ holds, teaching effort ($t^* = \hat{t} = \frac{2s(kF-2s^2)}{k^2-4s^2}$) and student enrollment ($n^* = \hat{n} = \frac{2s(2F-k)}{k^2-4s^2}$) are positively related to F, namely, $\frac{\partial t^*}{\partial F} = \frac{2ks}{k^2-4s^2} > 0$ and $\frac{\partial n^*}{\partial F} = \frac{4s}{k^2-4s^2} > 0$. When $\frac{2s^2}{k} < F \leq \frac{k}{2}$ and $F \geq \frac{k}{2} + \frac{k^2-4s^2}{4s}$, it is clear that $\frac{\partial t^*}{\partial F} = \frac{\partial n^*}{\partial F} = 0$.

(3) Since $\frac{\partial r^*}{\partial F} = \frac{k^2}{k^2-4s^2}$ and $\frac{\partial t^*}{\partial F} = \frac{2ks}{k^2-4s^2} > 0$ for $\frac{k}{2} < F < \frac{k}{2} + \frac{k^2-4s^2}{4s}$, we can derive $\frac{dr^*}{dt^*} = \frac{\partial r^*}{\partial F} / \frac{\partial t^*}{\partial F} = \frac{k}{2s} > 1$ under the assumption of $s < \frac{k}{2}$. ■

Proposition 3.4 (1) We can demonstrate that $\frac{\partial r^*}{\partial k} = \frac{2s^2(k^2-4kF+4s^2)}{(k^2-4s^2)^2} < \frac{2(\frac{k^2}{4})[k^2-4kF+4(\frac{k^2}{4})]}{(k^2-4s^2)^2} = k^3(k-2F) < 0$ because $s < \frac{k}{2}$ and $F > \frac{k}{2}$ in a positive interior equilibrium solution. Although $\frac{\partial t^*}{\partial k}$ can be directly calculated, we employ $t^* = (\frac{2s}{k})r^*$ that is suggested from the first-order condition of t. From this equation, we obtain $\frac{\partial t^*}{\partial k} = 2[\frac{ks(\frac{\partial r^*}{\partial k})-sr^*}{k^2}] < 0$. In addition, $n^* = \frac{2(t^*-s)}{k} > 0$ leads to $\frac{\partial n^*}{\partial k} = 2[\frac{k(\frac{\partial t^*}{\partial k})-(t^*-s)}{k^2}] < 0$ for $n^* > 0$, namely, $t^* > s$. Finally, we can derive $\frac{\partial R^*}{\partial k} = 2r^*(\frac{\partial r^*}{\partial k}) < 0$ because $R^* = (r^*)^2$.

(2) In the first place, $\frac{\partial r^*}{\partial s} = \frac{4k^2s(2F-k)}{(k^2-4s^2)^2} > 0$ can be demonstrated. From $t^* = (\frac{2s}{k})r^*$, we derive $\frac{\partial t^*}{\partial s} = \frac{2}{k}[r^* + s(\frac{\partial r^*}{\partial s})] > 0$. Next, note that the sign of $\frac{\partial n^*}{\partial s}$ depends on $\frac{\partial(t^*-s)}{\partial s}$. Thus, we obtain $\frac{\partial(t^*-s)}{\partial s} = (\frac{\partial t^*}{\partial s}) - 1 = \frac{2}{k}[r^* + s(\frac{\partial r^*}{\partial s})] - 1$. Since we have already derived $\frac{\partial r^*}{\partial s} > 0$, it can be shown that $\frac{\partial(t^*-s)}{\partial s} > r^*(\frac{2}{k}) - 1 = \frac{2r^*-k}{k}$. Examining the sign of the numerator

derives $2r^* - k = 2[\frac{k(kF-2s^2)}{k^2-4s^2}] - k = \frac{k^2(2F-k)}{k^2-4s^2} > 0$ for $F > \frac{k}{2}$. Hence, we can conclude $\frac{\partial n^*}{\partial s} > 0$. Lastly, $\frac{\partial R^*}{\partial s} = 2r^*(\frac{\partial r^*}{\partial s}) > 0$ is demonstrated. ■

Proposition 3.5 Because $r + t = \overline{a}$ holds at an equilibrium when $\hat{r} + \hat{t} > 0$, we can derive $t = \overline{a} - r$. By substituting it into the payoff function of the university, $U = r[F + \frac{2s(\overline{a}-r-s)}{k}] - \frac{r^2}{2} - \frac{(\overline{a}-r)^2}{2}$. The first-order condition $\frac{\partial U}{\partial r} = 0$ provides $r^* = \frac{kF-2s^2+\overline{a}(2s+k)}{2(2s+k)}$. By substituting r^* back into $t = \overline{a} - r$, we also obtain $t^* = \frac{-(kF-2s^2)+\overline{a}(2s+k)}{2(2s+k)}$. Hence, we can easily demonstrate that $\frac{\partial r^*}{\partial F} = \frac{k}{2(2s+k)} > 0$ and $\frac{\partial t^*}{\partial F} = -\frac{k}{2(2s+k)} < 0$. Since $n^* = \frac{2(t^*-s)}{k}$, we obtain $\frac{\partial n^*}{\partial F} = (\frac{2}{k})\frac{\partial t^*}{\partial F} = -\frac{1}{2s+k} < 0$. Furthermore, we can denote $R^* = r^*(F + sn^*)$, and thus, $\frac{\partial R^*}{\partial F} = \frac{\partial r^*}{\partial F}(F + sn^*) + r^*[1 + s(\frac{\partial n^*}{\partial F})]$. Because $1 + s(\frac{\partial n^*}{\partial F}) = 1 - \frac{s}{2s+k} = \frac{s+k}{2s+k} > 0$, we can see that the decrease in tuition revenue is smaller than an increase in the research fund. In sum, we conclude $\frac{\partial R^*}{\partial F} > 0$. ■

Proposition 3.6 (1) When $F < k$ holds, the saddle point $\tilde{e} = (k, \sqrt{2k(k-F)})$ in the diagram of (r, t) appears in the north-east space from the point, $(F, 0)$. Hence, the university finds it optimal to select either a maximum boundary effort, $\overline{e} = (\overline{r}, \overline{t})$ that satisfies both $r + t = \overline{a}$ and $r = F + \frac{t^2}{2k}$ (from Assumption 3.2), or a minimum boundary effort, $e^0 = (F + \delta_r, \delta_t)$ where δ_r and δ_t are infinitesimal positive values. First, if research and teaching effort is binding at $\overline{e} = (\overline{r}, \overline{t}) = (F + \frac{\overline{t}^2}{2k}, \overline{t})$, the payoff of the university reaches $U(F + \frac{\overline{t}^2}{2k}, \overline{t}) = \frac{1}{2}(F + \frac{\overline{t}^2}{2k})^2 - \frac{\overline{t}^2}{2} = \frac{\overline{t}^4+4k(F-k)\overline{t}^2+4k^2F^2}{8k^2}$. We need to compare this payoff with $U(F+\delta_r, \delta_t) \approx U(F, 0) = \frac{F^2}{2}$ at the minimum boundary effort. Solving the quadratic equation of $U(F + \frac{\overline{t}^2}{2k}, \overline{t}) = U(F, 0) \Leftrightarrow \frac{\overline{t}^4+4k(F-k)\overline{t}^2+4k^2F^2}{8k^2} = \frac{F^2}{2}$, we obtain $\overline{t}^2 = 4k(k-F) \Leftrightarrow \overline{t} = 2\sqrt{k(k-F)} > \sqrt{2k(k-F)}$ for $F < k$. From this relation, if $\overline{t} > 2\sqrt{k(k-F)}$ holds, the equilibrium solution is $e^* = \overline{e} = (F + \frac{\overline{t}^2}{2k}, \overline{t})$,

$s^* = \overline{s} = \frac{\bar{t}}{2}$, $n^* = \overline{n} = \frac{2(t^*-s^*)}{k} = \frac{\bar{t}}{k}$, and $R^* = r^*(F + s^*n^*) = (F + \frac{\bar{t}^2}{2k})^2$. By contrast, if $\bar{t} < 2\sqrt{k(k-F)}$ holds, the equilibrium solution is $\boldsymbol{e}^* = \boldsymbol{e}^0 = (F + \delta_r, \delta_t) \approx (F, 0)$, $s^* = \frac{\delta_t}{2} \approx 0$, $n^* = \frac{\delta_t}{k} \approx 0$, and $R^* = (F + \delta_r)(F + \frac{\delta_t^2}{2k}) \approx F^2$.

(2) As regards to $F > k$, while the saddle point $\tilde{\boldsymbol{e}} = (k, \sqrt{2k(k-F)})$ does not appear, the point $\boldsymbol{e}^0 = (F, 0)$ becomes the new saddle point. (See Subsection 3.8.2.) It is clear that $U(F + \frac{\bar{t}^2}{2k}, \bar{t}) = \frac{\bar{t}^4 + 4k(F-k)\bar{t}^2 + 4k^2F^2}{8k^2} > \frac{\bar{t}^2 + 4k^2F^2}{8k^2} = \frac{\bar{t}^2}{8k^2} + \frac{F^2}{2} > \frac{F^2}{2} = U(F, 0)$. Hence, the equilibrium solution is $\boldsymbol{e}^* = \overline{\boldsymbol{e}} = (F + \frac{\bar{t}^2}{2k}, \bar{t})$ for every $\bar{t}$. ■

Proposition 3.7 (1) Since $r^* + t^* = \bar{r} + \bar{t} = \overline{a}$ is satisfied as an equilibrium solution, $\boldsymbol{e}^* = \overline{\boldsymbol{e}} = (F + \frac{\bar{t}^2}{2k}, \bar{t})$, we obtain $r^* = F + \frac{\bar{t}^2}{2k} = F + \frac{(\overline{a}-r^*)^2}{2k}$. Taking a derivative with regard to F on both sides of the equation, we have $\frac{\partial r^*}{\partial F} = 1 - (\frac{\overline{a}-r^*}{k})\frac{\partial r^*}{\partial F}$, which can be transformed into $(1 + \frac{\overline{a}-r^*}{k})\frac{\partial r^*}{\partial F} = 1$. Since $1 + \frac{\overline{a}-r^*}{k}$ in the previous equation is obviously positive, we obtain $\frac{\partial r^*}{\partial F} > 0$. Moreover, since the capacity, $\overline{a}$, being constant indicates $\frac{\partial r^*}{\partial F} + \frac{\partial t^*}{\partial F} = 0$, we can conclude $\frac{\partial t^*}{\partial F} = -\frac{\partial r^*}{\partial F} < 0$. As for the other comparative statics, we can show that $\frac{\partial s^*}{\partial F} = (\frac{1}{2})\frac{\partial t^*}{\partial F} < 0$, $\frac{\partial n^*}{\partial F} = (\frac{1}{k})\frac{\partial t^*}{\partial F} < 0$, and $\frac{\partial R^*}{\partial F} = 2r^*(\frac{\partial r^*}{\partial F}) > 0$.

(2) The marginal increase in F moves the equilibrium solution only for research effort and research output, but not teaching effort and student enrollment, as suggested by the solution. Therefore, $\frac{\partial r^*}{\partial F} > 0$, $\frac{\partial t^*}{\partial F} = 0$, $\frac{\partial s^*}{\partial F} = 0$, $\frac{\partial n^*}{\partial F} = 0$, and $\frac{\partial R^*}{\partial F} > 0$ hold. ■

Proposition 3.8 (1) If we assume $\bar{t} > \hat{t} = \frac{2s(kF-2s^2)}{k^2-4s^2}$, $\bar{r} > \hat{r}$ is also expected to be satisfied because of $\hat{r} + \hat{t} < \overline{a}$ and $\bar{r} + \bar{t} = \overline{a}$ obtained from the construction.

(1-i) As shown in Proposition 3.6, when the capacity, $\overline{a}$, is large enough that $\bar{t} > 2\sqrt{k(k-F)}$ and $\frac{k}{2} < F < k$ are satisfied ($F > \frac{k}{2}$ is required for $\hat{n} > 0$), the university prefers $\overline{e} = (\overline{r}, \overline{t}) = (F + \frac{\overline{t}^2}{2k}, \overline{t})$ to $e^0 = (F + \delta_r, \delta_t) \approx (F, 0)$.

(1-ii) We have also proved that if $F > k$ is satisfied, the university always prefers $\overline{e} = (\overline{r}, \overline{t}) = (F + \frac{\overline{t}^2}{2k}, \overline{t})$ for every $\overline{t}$. In order for such an F to exist within $F \in (\frac{k}{2}, \frac{k}{2} + \frac{k^2-4s^2}{4s})$ (the condition of which is that $\hat{e} = (\hat{r}, \hat{t})$ is a positive interior equilibrium solution under a fixed tuition fee scheme), it must be the case that $\frac{k}{2} + \frac{k^2-4s^2}{4s} > k$. Hence, solving this quadratic inequality, we need to keep $s < \frac{k(\sqrt{5}-1)}{4} \approx 0.309k$. This satisfies the condition of $s < \frac{k}{2} = 0.5k$ that is necessary for the second-order condition.

(2) In order to check whether $\overline{n}$ is larger than $\hat{n}$, we examine whether $\overline{n} - \hat{n} = \frac{\overline{t}}{k} - \frac{2s(2F-k)}{k^2-4s^2} = \frac{(k^2-4s^2)\overline{t}-2ks(2F-k)}{k(k^2-4s^2)} > 0$ holds. By solving this inequality with regard to $\overline{t}$, we can demonstrate $\overline{n} > \hat{n}$ for $\overline{t} > \frac{2ks(2F-k)}{k^2-4s^2} = k\hat{n}$. Furthermore, $k\hat{n} - \hat{t} = \frac{2ks(2F-k)}{k^2-4s^2} - \frac{2s(kF-2s^2)}{k^2-4s^2} = \frac{2s(kF-2s^2)}{k^2-4s^2} > 0 \Leftrightarrow k\hat{n} > \hat{t}$ is satisfied. Hence, $\overline{t} > k\hat{n}$ is a stricter condition than $\overline{t} > \hat{t}$.

(3) As has been already shown, $R = r(F + sn) = r^2$ holds in this modeling from the first-order condition of r. When result (1) is applied, we obtain $\overline{R} = \overline{r}^2 > \hat{r}^2 = \hat{R}$.

(4) Let us denote $f(F) = 2\sqrt{k(k-F)}$ and $g(F) = \frac{2ks(2F-k)}{k^2-4s^2}$. When $\overline{t} > f(F)$ and $\overline{t} < g(F)$ hold, we can derive $\overline{R} > \hat{R}$ and $\overline{n} < \hat{n}$. Obviously, $f(F)$ is decreasing and $g(F)$ is increasing in F monotonically. We have $f(\frac{k}{2}) = \sqrt{2}k > 0$, $g(\frac{k}{2}) = 0$, $f(k) = 0$, $g(k) = \frac{2k^2s}{k^2-4s^2} > 0$, and $g(\frac{k}{2} + \frac{k^2-4s^2}{4s}) = k > 0$. Suppose that $s < \frac{k(\sqrt{5}-1)}{4} \approx 0.309k$ as before. As the diagram of Figure 3.10 illustrates, $f(F)$ and $g(F)$ must intersect only once at some point $F \in (\frac{k}{2}, k)$ from the intermediate-value theorem. Hence, we can clearly find that there exist $\overline{t} > f(F)$ and $\overline{t} < g(F)$ in the area of (A). ■

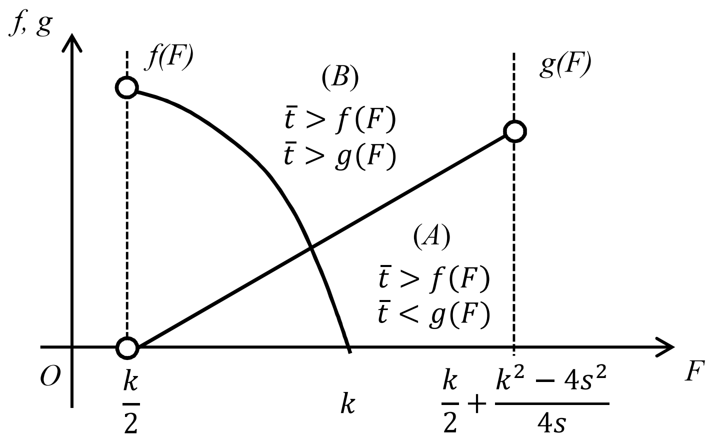


Figure 3.10. Diagram of $f(F)$ and $g(F)$.

Proposition 3.9 Consider the problem of university 1. Since the first-order condition $r_1 = \frac{F}{2} + \frac{s(t_1 - s)}{k} = b_1$ must be always satisfied, the payoff function is denoted by $U_1(r_1, t_1) = R_1 - (\frac{r_1^2}{2} + \frac{t_1^2}{2}) = \frac{r_1^2}{2} - \frac{t_1^2}{2} = \frac{1}{2}[\frac{F}{2} + \frac{s(t_1 - s)}{k}]^2 - \frac{t_1^2}{2}$ because of $R_1 = r_1 b_1 = r_1^2$. Now we try to seek t_1^{C1} that leads to $U_1 = 0$, where competition through teaching effort results in zero utilities of the universities. Thus, $(t_1^{C1})^2 = [\frac{F}{2} + \frac{s(t_1^{C1} - s)}{k}]^2$ holds, so that by solving by t_1^{C1}, we obtain $t_1^{C1} = \frac{kF - 2s^2}{2(k-s)}$. We also derive $r_1^{C1} = \frac{kF - 2s^2}{2(k-s)}$ by substituting $t_1^{C1} = \frac{kF - 2s^2}{2(k-s)}$ into the first-order condition of r. From these results, it can be demonstrated that $n_1^{C1} = \frac{t_1^{C1} - s}{k} = \frac{F - 2s}{2(k-s)}$ and $R_1^{C1} = (r_1^{C1})^2 = [\frac{kF - 2s^2}{2(k-s)}]^2$. (Note that although $F > 2s > \frac{2s^2}{k}$ is required for a positive interior equilibrium solution, the initial assumption of $F > 2k$ satisfies the above condition.) In addition, the total amount of each variable is provided by: $r^{C1} = r_1^{C1} + r_2^{C1} = \frac{kF - 2s^2}{k-s}$; $t^{C1} = t_1^{C1} + t_2^{C1} = \frac{kF - 2s^2}{k-s}$; $n^{C1} = n_1^{C1} + n_2^{C1} = \frac{F - 2s}{k-s}$; and $R^{C1} = R_1^{C1} + R_2^{C1} = \frac{1}{2}(\frac{kF - 2s^2}{k-s})^2$.

Based on these measures prepared, we compare the cases between single and multiple universities. (1) $r^{C1} - \hat{r} = \frac{kF - 2s^2}{k-s} - \frac{k(kF - 2s^2)}{k^2 - 4s^2} = \frac{s(k-4s)(kF - 2s^2)}{(k-s)(k^2 - 4s^2)} > 0$ for $s < \frac{k}{4} = 0.25k$. (2) $t^{C1} - \hat{t} = \frac{kF - 2s^2}{k-s} - \frac{2s(kF - 2s^2)}{k^2 - 4s^2} = \frac{(kF - 2s^2)(k^2 - 2ks - 2s^2)}{(k-s)(k^2 - 4s^2)} > 0$ for $s < \frac{(\sqrt{3} - 1)k}{2} \approx 0.366k$. (3)

$n^{C1} - \hat{n} = \frac{F-2s}{k-s} - \frac{2s(2F-k)}{k^2-4s^2} = \frac{(k-4s)(kF-2s^2)}{(k-s)(k^2-4s^2)} > 0$ for $s < \frac{k}{4} = 0.25k$. (4) $R^C - \hat{R} = \frac{1}{2}[\frac{kF-2s^2}{k-s}]^2 - [\frac{k(kF-2s^2)}{k^2-4s^2}]^2 = \frac{[(kF-2s^2)]^2[(k^2-4s^2)^2-2[k(k-s)]^2]}{2(k-s)^2(k^2-4s^2)^2}$. Now we can rewrite the second term of the nominator as $(k^2-4s^2)^2 - 2[k(k-s)]^2 = [k^2-4s^2+\sqrt{2}k(k-s)][k^2-4s^2-\sqrt{2}k(k-s)]$. Since the first term is strictly positive with $s < \frac{k}{2}$ (the second-order condition for a single university), we need to verify the sign of the second term. Let us denote $\eta(s) = k^2 - 4s^2 - \sqrt{2}k(k-s) = -4s^2 + \sqrt{2}ks + (1-\sqrt{2})k^2$. The fundamental discriminant of $\eta(s) = 0$ provides $(\sqrt{2}k)^2 + 16(1-\sqrt{2})k^2 = (18 - 16\sqrt{2})k^2 \approx -4.627k^2 < 0$. This suggests that $\eta(s) < 0$ for every s and k. Hence, we can conclude $R^{C1} < \hat{R}$ for every F, s, and k. ■

Proposition 3.10 (1-i) The government is assumed to regulate tuition fee to maximize tuition revenue (Assumption 3.3). From this, the tuition fee of university i ($i = 1, 2$) is set as $s_i = \frac{t_i}{2}$ by solving the problem, $\max_{s_i} E_i = \frac{s_i(t_i - s_i)}{k}$, based on the assumption that universities 1 and 2 evenly distribute student enrollment. Tuition revenue of university i is provided by $E_i = \frac{s_i(t_i-s_i)}{k} = \frac{t_i^2}{4k}$. Thus, the payoff function of university i can be represented as $U_i(r_i, t_i) = r_i(\frac{F}{2} + \frac{t_i^2}{4k}) - (\frac{r_i^2}{2} - \frac{t_i^2}{2})$. The first-order condition with regard to r_i is also satisfied from the assumption made (Assumption 3.2), so that we obtain $\frac{\partial U_i}{\partial r_i} = \frac{F}{2} - r_i + \frac{t_i^2}{4k} = 0 \Leftrightarrow r_i(t_i) = \frac{F}{2} + \frac{t_i^2}{4k}$. University i gains the payoff such that $U_i(r_i, t_i) = r_i(\frac{F}{2} + \frac{t_i^2}{4k}) - (\frac{r_i^2}{2} + \frac{t_i^2}{2}) = \frac{rF}{2} - \frac{r_i^2}{2} + (\frac{r_i - 2k}{4k})t_i^2$. Now let $(\bar{r}', \bar{t}')$ denote the research and teaching effort at the full capacity that satisfies $r_i + t_i = \bar{a}$ and $r_i = \frac{F}{2} + \frac{t_i^2}{4k}$. We investigate the point where $U_i(r_i, t_i) = U_i(\frac{F}{2}, 0)$ holds. Since $U_i(r_i, t_i) = \frac{r_i^2}{2} - \frac{t_i^2}{2} = \frac{1}{2}(\frac{F}{2} + \frac{t_i^2}{4k})^2 - \frac{t_i^2}{2} = \frac{t_i^4 + 4k(F-4k)t_i^2 + 4k^2F^2}{32k^2}$ and $U_i(\frac{F}{2}, 0) = \frac{F^2}{8}$ are derived, we obtain $t_i = 2\sqrt{k(4k-F)}$ by solving $\frac{t_i^4 + 4k(F-4k)t_i^2 + 4k^2F^2}{32k^2} = \frac{F^2}{8}$. Hence, if $\bar{t}' < 2\sqrt{k(4k-F)}$

holds, the university i is small enough to choose $(r_i, t_i) \approx (\frac{F}{2}, 0)$ in the absence of teaching competition. Consider first the case of $U_i(\bar{r}', \bar{t}') \geq 0$. Because competition only through teaching effort continues along with $r_i(t_i) = \frac{F}{2} + \frac{(t_i)^2}{4k}$ up to until the capacity is fully exploited, we obtain $\boldsymbol{e}_i^{C2} = \bar{\boldsymbol{e}}_i' = (\bar{r}', \bar{t}') = (\frac{F}{2} + \frac{(\bar{t}')^2}{4k}, \bar{t}')$. In addition, $s_i^{C2} = \frac{t_i^{C2}}{2} = \frac{\bar{t}'}{2}$, $n_i^{C2} = \frac{t_i^{C2} - s_1^{C2}}{k} = \frac{\bar{t}'}{2k}$, and $R_i^{C2} = (r_i^{C2})^2 = \frac{1}{4}[F + \frac{(\bar{t}')^2}{2k})]^2$ are derived.

(1-ii) When $U_i(\bar{r}', \bar{t}') < 0$ holds, we can demonstrate that there exists $\boldsymbol{e}_i^{C2} = \boldsymbol{e}' = (r', t')$ that satisfies $U_i(r', t') = 0$ on $r_i = \frac{F}{2} + \frac{t_i^2}{4k}$. The reason is that since $U_i(\frac{F}{2}, 0) = \frac{F^2}{8} > 0$ and $U_i(\bar{r}', \bar{t}') < 0$ holds, we can find some (r', t') that satisfies $\frac{F}{2} < r' < \bar{r}'$, $0 < t' < \bar{t}' < 2\sqrt{k(4k - F)}$, and $r' = \frac{F}{2} + \frac{(t')^2}{4k}$ from the intermediate-value theorem. We therefore obtain $\boldsymbol{e}_i^{C2} = \boldsymbol{e}' = (\frac{F}{2} + \frac{(t')^2}{4k}, t')$, $s_i^{C2} = \frac{t_i^{C2}}{2} = \frac{t'}{2}$, $n_i^{C2} = \frac{t_i^{C2} - s_1^{C2}}{k} = \frac{t'}{2k}$, and $R_i^{C2} = (r_i^{C2})^2 = \frac{1}{4}[F + \frac{(t')^2}{2k}]^2$.

(2-i) Universities 1 and 2 are symmetric in view of their capacities, so that $\boldsymbol{e}_1^{C2} = \boldsymbol{e}_2^{C2}$, $s_1^{C2} = s_2^{C2}$, $n_1^{C2} = n_2^{C2}$, and $R_1^{C2} = R_2^{C2}$ hold. In the case of (1-i), $r^{C2} = F + \frac{(\bar{t}')^2}{2k} > F = r^0$, $t^{C2} = 2\bar{t}' > 0 \approx t^0$, $n^{C2} = \frac{\bar{t}'}{k} > 0 \approx n^0$ are derived. In the case of (1-ii), $r^{C2} = F + \frac{(t')^2}{2k} > F = r^0$, $t^{C2} = 2t' > 0 \approx t^0$, $n^{C2} = \frac{t'}{k} > 0 \approx n^0$ can be similarly demonstrated.

(2-ii) Let us focus on the equilibrium solution in the case of (1-i). Solving the quadratic inequality, $R^{C2} > R^0 \Leftrightarrow \frac{1}{2}[F + \frac{(\bar{t}')^2}{2k}]^2 > F^2$, we have $\bar{t}' > \sqrt{2(\sqrt{2} - 1)kF}$. But since we have already assumed $\bar{t} < 2\sqrt{k(4k - F)}$ in the case of a single university which chooses minimum effort without competition, $\sqrt{2(\sqrt{2} - 1)kF} < 2\sqrt{k(4k - F)}$ is required in order for some $\bar{t}' \in \left(\sqrt{2(\sqrt{2} - 1)kF}, 2\sqrt{k(4k - F)}\right)$ to exist. We confirm whether such $\bar{t}'$ exists that satisfies the above inequality. In providing a comparison, we calculate $\left(2\sqrt{k(4k-F)}\right)^2 - \left(\sqrt{2(\sqrt{2}-1)kF}\right)^2 = 2k[8k-(1+\sqrt{2})F] > 2F[8F-(1+\sqrt{2})F] \approx 5.586F^2 > 0$ because of $F < k$. Hence, $\bar{t}'$ exists in the range, $\bar{t}' \in \left(\sqrt{2(\sqrt{2}-1)kF}, 2\sqrt{k(4k-F)}\right)$. If we take $\bar{t}' \to 2\sqrt{k(4k-F)}$, it is expected that the universities will choose $\bar{e}' = (\bar{r}', \bar{t}')$ since $U(\bar{r}', \bar{t}') > 0$ holds. From these derivations, we can conclude $R^{C2} > R^0$ when $\bar{t}'$ is sufficiently close to $2\sqrt{k(4k-F)}$. ■

3.8.2. Investigation of saddle points

In Subsection 3.5.2, the first-order conditions of maximizing $U(r,t) = Fr - \frac{r^2}{2} + (\frac{r-k}{2k})t^2$ with regard to r and t are given by Equations (3.35) and (3.36): $\frac{\partial U}{\partial r} = F - r + \frac{t^2}{2k} = 0$ and $\frac{\partial U}{\partial t} = (\frac{r-k}{k})t = 0$, respectively. By solving these two simultaneous equations, we derive the following solutions: $(r,t) = (k, \sqrt{2k(k-F)})$ and $(F, 0)$.

We define the Hessian matrix of $U(r,t)$ as follows:

$$\tilde{U} = \begin{bmatrix} \frac{\partial^2 U}{\partial r^2} & \frac{\partial^2 U}{\partial r \partial t} \\ \frac{\partial^2 U}{\partial t \partial r} & \frac{\partial^2 U}{\partial t^2} \end{bmatrix} = \begin{bmatrix} -1 & \frac{t}{k} \\ \frac{t}{k} & \frac{r-k}{k} \end{bmatrix}. \tag{3.40}$$

Let us first consider $(r,t) = (k, \sqrt{2k(k-F)})$ with $F < k$. In this case, $|\tilde{U}| = -\frac{t^2}{k} = -\frac{2k(k-F)}{k} < 0$, and hence, $(r,t) = (k, \sqrt{2k(k-F)})$ is a saddle point. On the other hand, evaluating at $(r,t) = (F,0)$, we obtain $|\tilde{U}| = -\frac{F-k}{k} > 0$ for $F < k$, which implies that $(r,t) = (F,0)$ is a local maximum. Next suppose $F > k$, then we derive only $(r,t) = (F,0)$ as a solution to the simultaneous equations. Since the determinant at this point is $|\tilde{U}| = -\frac{F-k}{k} < 0$, $(r,t) = (F,0)$ is a saddle point.

3.8.3. Comparative statics of other parameters

Student mobility cost (k)

Taking the derivatives on both sides of Equations (3.4) and (3.5) by k, respectively, we obtain the following relation:

$$s\left[\left.\frac{\partial n}{\partial k}\right|_{t=t_c}+\left(\frac{\partial n}{\partial t}\right)\frac{\partial t^*}{\partial k}\right]-\left[\left(\frac{\partial^2 C}{\partial r^2}\right)\frac{\partial r^*}{\partial k}+\left(\frac{\partial^2 C}{\partial r\partial t}\right)\frac{\partial t^*}{\partial k}\right]=0$$
$$\Longleftrightarrow\left(\frac{\partial^2 C}{\partial r^2}\right)\frac{\partial r^*}{\partial k}+\left[\frac{\partial^2 C}{\partial r\partial t}-s\left(\frac{\partial n}{\partial t}\right)\right]\frac{\partial t^*}{\partial k}=s\left(\left.\frac{\partial n}{\partial k}\right|_{t=t_c}\right), \tag{3.41}$$
$$s\left[\left(\frac{\partial n}{\partial t}\right)\frac{\partial r^*}{\partial k}+r^*\left[\frac{\partial^2 n}{\partial t\partial k}+\left(\frac{\partial^2 n}{\partial t^2}\right)\frac{\partial t^*}{\partial k}\right]\right]-\left[\left(\frac{\partial^2 C}{\partial r\partial t}\right)\frac{\partial r^*}{\partial k}+\left(\frac{\partial^2 C}{\partial t^2}\right)\frac{\partial t^*}{\partial k}\right]=0$$
$$\Longleftrightarrow\left[\frac{\partial^2 C}{\partial r\partial t}-s\left(\frac{\partial n}{\partial t}\right)\right]\frac{\partial r^*}{\partial k}+\left[\frac{\partial^2 C}{\partial t^2}-sr^*\left(\frac{\partial^2 n}{\partial t^2}\right)\right]\frac{\partial t^*}{\partial k}=sr^*\left(\frac{\partial^2 n}{\partial t\partial k}\right), \tag{3.42}$$

where $t=t_c$ means it is a derivative on the condition that t is constant. The matrix notation of Equations (3.41) and (3.42) is:

$$\begin{bmatrix}\frac{\partial^2 C}{\partial r^2} & \frac{\partial^2 C}{\partial r\partial t}-s(\frac{\partial n}{\partial t})\\ \frac{\partial^2 C}{\partial r\partial t}-s(\frac{\partial n}{\partial t}) & \frac{\partial^2 C}{\partial t^2}-sr^*(\frac{\partial^2 n}{\partial t^2})\end{bmatrix}\begin{bmatrix}\frac{\partial r^*}{\partial k}\\ \frac{\partial t^*}{\partial k}\end{bmatrix}=\begin{bmatrix}s(\left.\frac{\partial n}{\partial k}\right|_{t=t_c})\\ sr^*(\frac{\partial^2 n}{\partial t\partial k})\end{bmatrix}. \tag{3.43}$$

Let us denote the first matrix as A_k. Suppose that $|A_k|=(\frac{\partial^2 C}{\partial r^2})[\frac{\partial^2 C}{\partial t^2}-sr^*(\frac{\partial^2 n}{\partial t^2})]-[\frac{\partial^2 C}{\partial r\partial t}-s(\frac{\partial n}{\partial t})]^2>0$ is satisfied for an analytical purpose. [37]

We solve Equation (3.43) with regard to $\frac{\partial r^*}{\partial k}$ and $\frac{\partial t^*}{\partial k}$:

$$\frac{\partial r^*}{\partial k}=\frac{\begin{vmatrix}s(\left.\frac{\partial n}{\partial k}\right|_{t=t_c}) & \frac{\partial^2 C}{\partial r\partial t}-s(\frac{\partial n}{\partial t})\\ sr^*(\frac{\partial^2 n}{\partial t\partial k}) & \frac{\partial^2 C}{\partial t^2}-sr^*(\frac{\partial^2 n}{\partial t^2})\end{vmatrix}}{|A_k|}$$

[37] As mentioned in Footnote 15, when $\frac{\partial^2 C}{\partial r\partial t}>0$ is assumed, $|A_k|$ is always satisfied from the assumption made with $C(r,t)$ and $n(t,k,s)$.

$$= \frac{1}{|A_k|}\left[s\left(\frac{\partial n}{\partial k}\bigg|_{t=t_c}\right)\left[\left(\frac{\partial^2 C}{\partial t^2}\right) - sr^*\left(\frac{\partial^2 n}{\partial t^2}\right)\right] - sr^*\left(\frac{\partial^2 n}{\partial t \partial k}\right)\left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\right], \quad (3.44)$$

$$\frac{\partial t^*}{\partial k} = \frac{\begin{vmatrix} \frac{\partial^2 C}{\partial r^2} & s(\frac{\partial n}{\partial k}\big|_{t=t_c}) \\ \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) & sr^*(\frac{\partial^2 n}{\partial t \partial k}) \end{vmatrix}}{|A_k|}$$

$$= \frac{1}{|A_k|}\left[sr^*\left(\frac{\partial^2 C}{\partial r^2}\right)\left(\frac{\partial^2 n}{\partial t \partial k}\right) - s\left(\frac{\partial n}{\partial k}\bigg|_{t=t_c}\right)\left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\right]. \quad (3.45)$$

In Equations (3.44) and (3.45), the signs of $\frac{\partial r^*}{\partial k}$ and $\frac{\partial r^*}{\partial k}$ are not decisive depending on the substitutability, $\frac{\partial^2 C}{\partial r \partial t}$. As usually expected, the conditions for the negative impact of a mobility cost on research and teaching efforts, $\frac{\partial r^*}{\partial k} < 0$ and $\frac{\partial r^*}{\partial k} < 0$, are as follows:

$$\frac{\partial r^*}{\partial k} < 0 \Longleftrightarrow \frac{\partial^2 C}{\partial r \partial t} < s\underbrace{\left(\frac{\partial n}{\partial t}\right)}_{\text{positive}} + \underbrace{\frac{(\frac{\partial n}{\partial k}\big|_{t=t_c})[(\frac{\partial^2 C}{\partial t^2}) - sr^*(\frac{\partial^2 n}{\partial t^2})]}{r^*(\frac{\partial^2 n}{\partial t \partial k})}}_{\text{positive}}, \quad (3.46)$$

$$\frac{\partial t^*}{\partial k} < 0 \Longleftrightarrow \frac{\partial^2 C}{\partial r \partial t} < s\underbrace{\left(\frac{\partial n}{\partial t}\right)}_{\text{positive}} + \underbrace{\frac{r^*(\frac{\partial^2 C}{\partial r^2})(\frac{\partial^2 n}{\partial t \partial k})}{\frac{\partial n}{\partial k}\big|_{t=t_c}}}_{\text{positive}}. \quad (3.47)$$

Equations (3.46) and (3.47) suggest that if research and teaching activities are complementary or independent (i.e. $\frac{\partial^2 C}{\partial r \partial t} \leq 0$), a rise in a mobility cost decreases both research and teaching effort. Consequently, we can also obtain $\frac{\partial n^*}{\partial k} = \frac{\partial n^*}{\partial k}\big|_{t=t_c} + (\frac{\partial n}{\partial t})(\frac{\partial t^*}{\partial k}) < 0$ and $\frac{\partial R^*}{\partial k} = \frac{\partial r^*}{\partial k}(F + sn^*) + r^* s(\frac{\partial n^*}{\partial k}) < 0$, which implies that both student enrollment and research output will decrease.

On the contrary, if these conditions are not satisfied (substitutability is sufficiently positive), we could obtain $\frac{\partial r^*}{\partial k} > 0$ and $\frac{\partial t^*}{\partial k} > 0$. In addition, substitutability being very strong could provide $\frac{\partial n^*}{\partial k} = \frac{\partial n^*}{\partial k}\big|_{t=t_c} + (\frac{\partial n^*}{\partial t})(\frac{\partial t^*}{\partial k}) > 0$ and $\frac{\partial R^*}{\partial k} = \frac{\partial r^*}{\partial k}(F + sn^*) + r^* s(\frac{\partial n^*}{\partial k}) > 0$. Although this argument seems somewhat surprising, the intuition is straightforward.

That is, a rise in a mobility cost in the first place reduces the budget of the university through a decrease in student enrollment, so that the university relinquishes some degree of research effort. However, if substitutability is strong enough, teaching effort in turn increases in response to the decrease in research effort, which culminates in higher student enrollment at the end. When this latter positive effect on student enrollment is sufficiently large, research effort and research output could be increased due to an enriched research budget.

Tuition fee (s)

A tuition fee is assumed to be an exogenous variable. If we take the derivatives on both sides of Equations (3.4) and (3.5) by s, respectively, we obtain:

$$\begin{aligned}
& n^* + s\left[\left.\frac{\partial n}{\partial s}\right|_{t=t_c} + \left(\frac{\partial n}{\partial t}\right)\frac{\partial t^*}{\partial s}\right] - \left[\left(\frac{\partial^2 C}{\partial r^2}\right)\frac{\partial r^*}{\partial s} + \left(\frac{\partial^2 C}{\partial r \partial t}\right)\frac{\partial t^*}{\partial s}\right] = 0 \\
& \Longleftrightarrow \left(\frac{\partial^2 C}{\partial r^2}\right)\frac{\partial r^*}{\partial s} + \left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\frac{\partial t}{\partial s} = n^* + s\left(\left.\frac{\partial n}{\partial s}\right|_{t=t_c}\right), \qquad (3.48) \\
& \left[s\left(\frac{\partial n}{\partial t}\right)\frac{\partial r^*}{\partial s} + r^*\left(\frac{\partial n}{\partial t}\right)\right] + r^* s\left[\frac{\partial^2 n}{\partial t \partial s} + \left(\frac{\partial^2 n}{\partial t^2}\right)\frac{\partial t^*}{\partial s}\right] - \left[\left(\frac{\partial^2 C}{\partial r \partial t}\right)\frac{\partial r^*}{\partial s} + \left(\frac{\partial^2 C}{\partial t^2}\right)\frac{\partial t^*}{\partial s}\right] = 0 \\
& \Longleftrightarrow \left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\frac{\partial r^*}{\partial s} + \left[\left(\frac{\partial^2 C}{\partial t^2}\right) - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right]\frac{\partial t^*}{\partial s} = r^*\left[\frac{\partial n}{\partial t} + s\left(\frac{\partial^2 n}{\partial t \partial s}\right)\right]. \qquad (3.49)
\end{aligned}$$

From Equations (3.48) and (3.49), the following matrix notation is derived:

$$\begin{bmatrix} \frac{\partial^2 C}{\partial r^2} & \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) \\ \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) & \frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2}) \end{bmatrix} \begin{bmatrix} \frac{\partial r}{\partial s} \\ \frac{\partial t}{\partial s} \end{bmatrix} = \begin{bmatrix} n^* + s(\left.\frac{\partial n}{\partial s}\right|_{t=t_c}) \\ r^*[\frac{\partial n}{\partial t} + s(\frac{\partial^2 n}{\partial t \partial s})] \end{bmatrix}. \qquad (3.50)$$

Suppose again that the determinant of the first matrix, $|A_s| = (\frac{\partial^2 C}{\partial r^2})[\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})] - [\frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t})]^2 > 0$. We obtain $\frac{\partial r^*}{\partial s}$ and $\frac{\partial t^*}{\partial s}$ such that:

$$\frac{\partial r^*}{\partial s} = \frac{\begin{vmatrix} n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c}) & \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) \\ r^*[\frac{\partial n}{\partial t} + s(\frac{\partial^2 n}{\partial t \partial s})] & \frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2}) \end{vmatrix}}{|A|}$$
$$= \frac{1}{|A_s|}\left[\left[\frac{\partial^2 C}{\partial t^2} - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right]\left[n^* + s\left(\frac{\partial n}{\partial s}\bigg|_{t=t_c}\right)\right] - r^*\left[\frac{\partial n}{\partial t} + s\left(\frac{\partial^2 n}{\partial t \partial s}\right)\right]\left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\right], \tag{3.51}$$

$$\frac{\partial t^*}{\partial s} = \frac{\begin{vmatrix} \frac{\partial^2 C}{\partial r^2} & n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c}) \\ \frac{\partial^2 C}{\partial r \partial t} - s(\frac{\partial n}{\partial t}) & r^*[\frac{\partial n}{\partial t} + s(\frac{\partial^2 n}{\partial t \partial s})] \end{vmatrix}}{|A|}$$
$$= \frac{1}{|A_s|}\left[r^*\left(\frac{\partial^2 C}{\partial r^2}\right)\left[\frac{\partial n}{\partial t} + s\left(\frac{\partial^2 n}{\partial t \partial s}\right)\right] - \left[n^* + s\left(\frac{\partial n}{\partial s}\bigg|_{t=t_c}\right)\right]\left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\right]. \tag{3.52}$$

Equations (3.51) and (3.52) reveal that the signs of $\frac{\partial r^*}{\partial s}$ and $\frac{\partial t^*}{\partial s}$ are dependent on those of $\frac{\partial^2 C}{\partial r \partial t}$ (the degree of substitutability), $n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c})$ (the degree of the tuition fee elasticity of student enrollment given teaching effort), [38] and $\frac{\partial^2 n}{\partial t \partial s}$ (the second-order differential coefficient with regard to teaching effort and the tuition fee). For analytical simplicity, we posit $\frac{\partial^2 n}{\partial t \partial s} = 0$, which implies that teaching effort and the tuition fee are independent in the student enrollment function, $n(t, k, s)$. By doing so, we can rewrite Equations (3.51) and (3.52) as follows:

$$\frac{\partial r^*}{\partial s} = \frac{1}{|A_s|}\left[\left[\frac{\partial^2 C}{\partial t^2} - r^* s\left(\frac{\partial^2 n}{\partial t^2}\right)\right]\left[n^* + s\left(\frac{\partial n}{\partial s}\bigg|_{t=t_c}\right)\right] - r^*\left(\frac{\partial n}{\partial t}\right)\left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\right], \tag{3.53}$$
$$\frac{\partial t^*}{\partial s} = \frac{1}{|A_s|}\left[r^*\left(\frac{\partial n}{\partial t}\right)\left(\frac{\partial^2 C}{\partial r^2}\right) - \left[n^* + s\left(\frac{\partial n}{\partial s}\bigg|_{t=t_c}\right)\right]\left[\frac{\partial^2 C}{\partial r \partial t} - s\left(\frac{\partial n}{\partial t}\right)\right]\right]. \tag{3.54}$$

[38] The sign of $n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c})$ being positive (negative) means that the elasticity is inelastic (elastic).

Thus, the conditions for $\frac{\partial r^*}{\partial s} < 0$ and $\frac{\partial t^*}{\partial s} < 0$ are derived as follows:

$$\frac{\partial r^*}{\partial s} < 0 \Longleftrightarrow \frac{\partial^2 C}{\partial r \partial t} > s \underbrace{\left(\frac{\partial n}{\partial t}\right)}_{\text{positive}} + \frac{[\frac{\partial^2 C}{\partial t^2} - r^* s(\frac{\partial^2 n}{\partial t^2})][n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c})]}{r^*(\frac{\partial n}{\partial t})}, \tag{3.55}$$

$$\frac{\partial t^*}{\partial s} < 0 \Longleftrightarrow \frac{\partial^2 C}{\partial r \partial t} > s \underbrace{\left(\frac{\partial n}{\partial t}\right)}_{\text{positive}} + \underbrace{\frac{r^*(\frac{\partial^2 C}{\partial r^2})(\frac{\partial n}{\partial t})}{n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c})}}_{\text{positive}} > 0 \quad \text{for } n^* + s\left(\frac{\partial n}{\partial s}\bigg|_{t=t_c}\right) > 0, \tag{3.56}$$

$$\Longleftrightarrow \frac{\partial^2 C}{\partial r \partial t} < s \underbrace{\left(\frac{\partial n}{\partial t}\right)}_{\text{positive}} + \underbrace{\frac{r^*(\frac{\partial^2 C}{\partial r^2})(\frac{\partial n}{\partial t})}{n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c})}}_{\text{negative}} \quad \text{for } n^* + s\left(\frac{\partial n}{\partial s}\bigg|_{t=t_c}\right) < 0, \tag{3.57}$$

The comparative statics of a tuition fee is much more complicated than a mobility cost. Equation (3.55) indicates that when certain degree of substitutability occurs (the right-hand side of the equation can be positive or negative), research effort is decreased, i.e. $\frac{\partial r^*}{\partial s} < 0$, by a rise in the tuition fee although the total budget may increase. On the other hand, there are two cases for $\frac{\partial t^*}{\partial s} < 0$ in accordance with the sign of $n^* + s(\frac{\partial n}{\partial s}\big|_{t=t_c})$. Let us first focus on Equation (3.56). When the tuition fee elasticity of student enrollment is inelastic given teaching effort, a rise in the tuition fee increases tuition revenue of the university, and research effort increases accordingly. There is also more room for enhancing teaching activities due to an increased university budget. But if substitutability between research and teaching activities is strong, teaching effort is decreased in an equilibrium. With regard to Equation (3.57), since the tuition fee elasticity is elastic, the tuition fee rise reduces tuition revenue. The university is compelled to reduce research effort. Whereat, if the degree of substitutability is not large, teaching effort is also ultimately decreased in an equilibrium. Both in the two cases of Equations (3.56) and (3.57), if teaching effort is finally reduced, student enrollment also decreases. Moreover, as for research output, $\frac{\partial R^*}{\partial s} = \frac{\partial r^*}{\partial s}(F + sn^*) + r^*[n^* + s[\frac{\partial n}{\partial s}\big|_{t=t_c} + (\frac{\partial n}{\partial t})(\frac{\partial t^*}{\partial s})]] < 0$ could be derived when the degree of substitutability is large and the tuition fee elasticity is inelastic. (Note that both $\frac{\partial r^*}{\partial s} < 0$ and $\frac{\partial t^*}{\partial s} < 0$ hold in this situation.) Lastly, we can obtain the conditions for $\frac{\partial r^*}{\partial s} > 0$ and $\frac{\partial t^*}{\partial s} > 0$ by reversing inequality signs of Equations (3.55)-(3.57), and the essence of reasoning is the same as the previous discussion.

3.8.4. Increasing returns to scale in a research output function

Consider the payoff function of a university, $U(r, t) = R - C(r, t)$, where the research output function is provided by $R = rb = r(F + sn)$. The first-order condition with regard to r is $F + sn - \frac{\partial C}{\partial r} = 0 \Leftrightarrow b = \frac{\partial C}{\partial r}$. Accordingly, the research output function can be represented as $R = rb = r(\frac{\partial C}{\partial r})$. By taking a derivative of R by r, we obtain:

$$\frac{\partial R}{\partial r} = \frac{\partial C}{\partial r} + r\left(\frac{\partial^2 C}{\partial r^2}\right) > 0, \tag{3.58}$$

$$\frac{\partial^2 R}{\partial r^2} = 2\left(\frac{\partial^2 C}{\partial r^2}\right) + r\left(\frac{\partial^3 C}{\partial r^3}\right). \tag{3.59}$$

Whether $\frac{\partial^2 R}{\partial r^2}$ is positive (increasing returns to scale) or negative (diminishing returns to scale) depends on, $\frac{\partial^3 C}{\partial r^3}$, which is the third derivative of the cost function with regard to research effort. As is easily shown, $\frac{\partial^3 C}{\partial r^3} \geq 0$ is a sufficient condition for $\frac{\partial^2 R}{\partial r^2} > 0$. But if an assumption is made that the order of $C(r, t)$ with regard to r is no less than 2, $\frac{\partial^2 R}{\partial r^2} > 0$ is always attainable. (Our illustrative model assumes $\frac{\partial^3 C}{\partial r^3} = 0$.) It is therefore concluded that the case, where the result of increasing returns to scale in a research output function collapses, seems to be limited.

3.8.5. Detailed explanation of an empirical analysis

Dataset

The primary indicators used in this empirical analysis are obtained from statistical data on the U.S. science and engineering institutions, which is supplied by the *National Center for Science Engineering Statistics (NCSES)*. The *NCSES*, which is under the supervision of the National Science Foundation (NSF), is responsible for compiling an assortment of U.S. competitiveness data in the field of R&D, science, engineering, and technology, as well as science-related education. For the use of this study, the variables of the number of doctorates awarded, student enrollment, publicly funded R&D, tuition fee, and the number of academic staff are collected from *NSF Survey of Earned Doctorates/Doctorate Records File*, *IPEDS Enrollment Survey*, *NSF Survey of Research and Development Expenditures at Universities and Colleges/Higher Education Research and Development Survey*, *IPEDS Institutional Characteristics Survey Tuition Data*, and *IPEDS Salaries, Tenure, and Fringe Benefits Survey*, respectively. These datasets are collectively provided on the WebCASPAR database. [39]

The object of this study is the 956 U.S. universities and colleges (hereafter, simply denoted “universities") that have recorded R&D expenditures at least in one year from 2003 to 2011. Note that the collected data of some universities is commensurate with the campus level. Among these 956 universities, 920 produce doctorate degrees and 905 obtain a positive value of student enrollment as of 2011. The universities are

[39] WebCASPAR is an integrated resources data system and provides easy access to statistical data of science and engineering at U.S. academic institutions. This database is available at: https://ncsesdata.nsf.gov/webcaspar/

comprised of 517 public and 439 private universities, respectively, and the estimation is conducted according to these university categories. The period subject to estimation is the years between 2003 and 2011. If there are some missing values in this period, such samples are totally deleted in regressions, so that the samples used for regressions are smaller than the original total.

In a series of estimations, the number of doctorates awarded (“total" and “natural science") and total student enrollment (including undergraduates and graduates) are used as dependent variables. Doctorate degrees awarded are assumed to represent approximate research output. Although, for example, patent counts and published articles are generally viewed as the appropriate indicators of research output, it would be difficult to collect such data for every university from our dataset. Since doctorate degrees are usually awarded in consideration of certain original works, they are likely to be related to research output produced by universities. Here both “total" and “natural science" doctorate degrees are employed. While the former consists of all academic disciplines including social sciences, the latter is confined to so-called natural sciences, such as engineering, physical science, geometry, geoscience, mathematics, life science, psychology, architecture, and these inter-disciplinary sciences.

Student enrollment denotes a head count of students that register with universities at the beginning of the fall term in each year. Total student enrollment is comprised of both undergraduate and graduate students. It should be noted that although the dataset mostly includes positive values of student enrollment, a small number of observations have a problem of missing values in this dataset. Since we simply omit the samples that include missing values instead of assigning them with zeros, the final samples do not entail zero values of student enrollment in conducting the regression analysis. As a result, the observations are different from those of the number of doctorates awarded that include “zeros" in the samples.

With regard to independent variables, this study utilizes not only federally funded general R&D (i.e. general R&D) but also federally and locally (state-governmentally) funded science and engineering R&D (i.e. science R&D) as indicators of public research funds. Strictly speaking, although this type of publicly funded R&D does not necessarily coincide with the research funds themselves allocated by government authorities, it can be regarded as a rough approximation to them. Tuition fees are divided into the categories of undergraduate and graduate. In the original data, they are listed by “In-State" and “Out-of-State" tuition fees (“In-State tuition" fees are generally more inexpensive than “Out-of-State" tuition fees), but the tuition fees used for estimation are transformed into the simple average of the two. In order to control for the capacities of universities, this analysis incorporates the number of academic staff as an independent variable. [40] Finally, the amount of R&D and the tuition fees are deflated across all time dimensions in reference to the implicit price deflater that is based on the fiscal year GDP from 2005 to June 2010. (The deflater is normalized to 1 in 2005.)

The descriptive statistics of these dependent and independent variables such as the

[40] Although the number of academic staff can be interpreted as an indication of research and teaching effort of universities, it appears to be significantly correlated with the capacity limitation. In other words, it is impossible to conduct more activities than the existing academic staff can manage.

mean and the standard deviation across all observations from 2003 to 2011 are exhibited for all, public, and private universities in Table 3.1. We can see that there are large differences in the variables between public and private universities except general R&D. More precisely, the scales of public universities represented by the number of doctorates awarded, students, and academic staff are larger and their tuition fees are less expensive than private universities, possibly due to financial assistance provided by federal and/or local governments. Hence, it is expected that some estimation results presented below give rise to the different features between these two university categories.

	All	Public	Private	*t*-value	*p*-value
Year: 2003-2011					
Total number of universities	956	517	439		
Simple average across all observations					
Dependent variables					
Number of doctorates awarded	48.7	63.6	31.1	12.554	0.000
(Total)	(118.9)	(136.7)	(90.7)		
Number of doctorates awarded	29.7	39.1	18.6	12.251	0.000
(Science)	(76.7)	(88.6)	(57.8)		
Total student enrollment	9513.3	13400.2	4897.8	41.709	0.000
	(10096.6)	(11255.2)	(5773.8)		
Independent variables					
General R&D ($ mil.)	41.628	42.571	40.295	0.860	0.390
	(103.007)	(86.767)	(122.322)		
Science R&D ($ mil.)	44.605	47.301	40.800	2.357	0.019
	(107.329)	(94.262)	(123.362)		
Undergraduate tuition fee ($ thou.)	14.860	8.069	23.432	110.391	0.000
	(0.111)	(0.048)	(0.144)		
Graduate tuition fee ($ thou.)	12.053	8.639	16.972	54.753	0.000
	(7.466)	(3.498)	(8.801)		
Number of academic staff	400.8	516.1	261.2	28.952	0.000
	(411.8)	(455.9)	(296.2)		

Note: 1. The *t* values represent statistics for the test of the difference in average between public and private universities.
2. The standard deviations are reported in the round parentheses.
3. Some samples are omitted due to missing values.

Table 3.1. Descriptive statistics of variables.

As we can observe from Figure 3.11, the distribution of the total number of doctorates awarded in 2011 has many “zeros" across the board and an elongated right-hand tail. In addition, upon examining Table 3.2 that indicates the relationship between the total number of doctorates awarded and academic staff in 2011, we find many small combinations such as “zero doctorates awarded and few academic staff", in particular, for private universities. Indeed, 45.8% and 61.0% of public and private universities, respectively, whose number of academic staff is less than 500, produce zero doctorates. [41] Moreover, Table 3.3 presents that student enrollment also tends to be small if the number of academic staff is few. The combination of less than 5000 student enrollment and less than 500 academic staff accounts for 22.6% and 66.3% for public and private universities, respectively. In fact, these findings are not entirely consistent with the theoretical result derived in Proposition 3.6, which strictly indicates that while research output produced by a university with a small capacity is still positive (the smallest, though), student enrollment is nearly zero. And yet, it is important to note these distinctive characteristics of the distributions, especially when we run a regression of the number of (science) doctorates awarded.

[41] We can also confirm that the number of science doctorates awarded exhibits the same tendency as along with the total number of science doctorates awarded, although the detailed graph and table are omitted to save space.

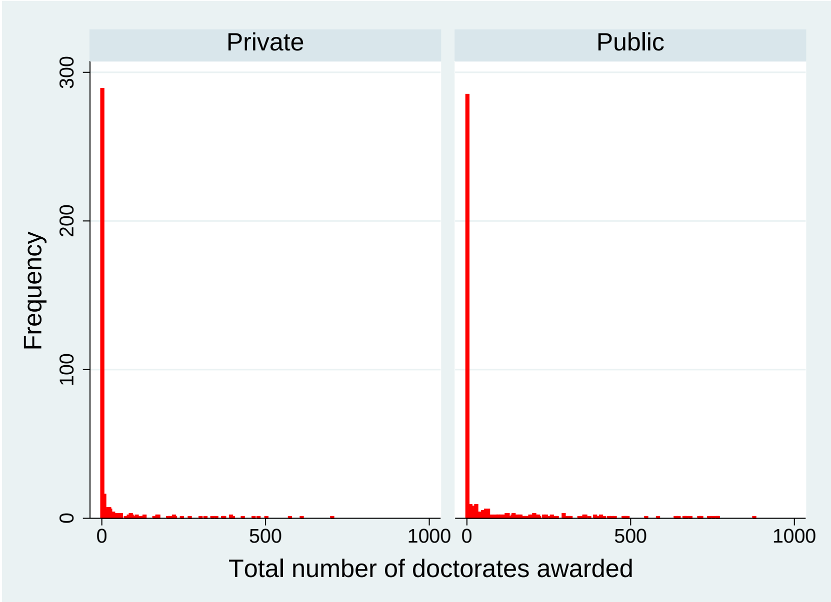


Figure 3.11. Frequency of the total number of doctorates awarded in 2011.

Public universities

	Total #doctorates awarded				
	= 0	1—49	50—99	100—149	≥ 150
#Acad. staff					
< 500	230 (45.8%)	50 (10.0%)	14 (2.8%)	2 (0.4%)	3 (0.6%)
500—999	36 (7.2%)	31 (6.2%)	19 (3.8%)	18 (3.6%)	13 (2.6%)
1000—1499	1 (0.2%)	3 (0.6%)	0 (0.0%)	4 (0.8%)	39 (7.8%)
1500—1999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	13 (2.6%)
2000—2499	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	9 (1.8%)
2500—2999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)
≥ 3000	10 (2.0%)	4 (0.8%)	1 (0.2%)	0 (0.0%)	1 (0.2%)

Private universities

	Total #doctorates awarded				
	= 0	1—49	50—99	100—149	≥ 150
#Acad. staff					
< 500	263 (61.0%)	64 (14.8%)	10 (2.3%)	2 (0.5%)	11 (2.6%)
500—999	2 (0.5%)	9 (2.1%)	8 (1.9%)	5 (1.2%)	13 (3.0%)
1000—1499	0 (0.0%)	0 (0.0%)	1 (0.2%)	1 (0.2%)	9 (2.1%)
1500—2999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (1.2%)
2000—2499	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)
2500—2999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
≥ 3000	21 (4.9%)	6 (1.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Table 3.2. Total number of doctorates awarded and academic staff in 2011.

Public universities

	Total student enrollment				
	< 5000	5000—9999	10000—14999	15000—19999	≥ 20000
#Acad. staff					
< 500	111 (22.3%)	119 (24.2%)	58 (11.8%)	9 (1.8%)	2 (0.4%)
500—999	0 (0.0%)	1 (0.2%)	21 (4.3%)	41 (8.3%)	54 (11.0%)
1000—1499	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)	46 (9.3%)
1500—1999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	13 (2.6%)
2000—2499	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	9 (1.8%)
2500—2999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)
≥ 3000	5 (1.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)

Private universities

	Total student enrollment				
	< 5000	5000—9999	10000—14999	15000—19999	≥ 20000
#Acad. staff					
< 500	275 (66.3%)	61 (14.7%)	10 (2.4%)	2 (0.5%)	2 (0.5%)
500—999	0 (0.0%)	8 (1.9%)	17 (4.1%)	6 (1.4%)	6 (1.4%)
1000—1499	0 (0.0%)	0 (0.0%)	4 (1.0%)	2 (0.5%)	5 (1.2%)
1500—1999	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)	4 (1.0%)
2000—2499	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	9 (2.2%)
2500—2999	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
≥ 3000	11 (2.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Table 3.3. Total student enrollment and academic staff in 2011.

Estimation methods

When the total number of doctorates awarded is a dependent variable, we need to recall the distribution of this variable that was shown in Figure 3.11, which exhibits the observations filled with many zeros and discrete values. Since a linear model is not applicable to this distribution, it is appropriate to use a fixed-effects Poisson (FEP) model for count data developed by Wooldridge (1999). In reference to Greene (2011), the FEP regression for panel data is defined as follows:

$$p\left(y_{i1}, y_{i2}, \cdots, y_{iT_i} \Big| \sum_{t=1}^{T_i} y_{it}\right) = \frac{(\sum_{t=1}^{T_i} y_{it})!}{(\prod_{t=1}^{T_i} y_{it}!)} \prod_{t=1}^{T_i} p_{it}^{y_{it}},$$
$$\text{where} \quad p_{it} = \frac{e^{\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i}}{\sum_{t=1}^{T_i} e^{\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i}} = \frac{e^{\boldsymbol{x}_{it}'\boldsymbol{\beta}}}{\sum_{t=1}^{T_i} e^{\boldsymbol{x}_{it}'\boldsymbol{\beta}}}. \tag{3.60}$$

An estimation, which is not dependent on the fixed effects anymore, can be derived by obtaining the joint distribution of $(y_{i1}, y_{i2}, \cdots, y_{iT_i})$ conditional on their sum (Greene, 2011). y_{it} and vector $\boldsymbol{x}_{it}$ denote dependent and independent variables for individual i at time t, respectively. Vector $\boldsymbol{\beta}$ represents regression coefficients, and α_i is an individual fixed effect. The contribution of individual i to the conditional likelihood is simply calculated as: $\ln L_i = \sum_{t=1}^{T_i} y_{it} \ln p_{it}$. In conducting estimations, our inference is based upon the robust clustered standard errors.

The FEP estimator, allowing for the dependence between the fixed effects intrinsic to universities and independent variables, is strongly robust for consistency under the conditional mean assumption. [42] On the other hand, we should take note that there could be a reverse causality that operates from the number of doctorates awarded to publicly funded R&D. That is, since it could be that an increase in the number of doctoral students attracts more research funds, we need to be cautious about judging the rigorous direction of causality (i.e. simultaneity bias). [43]

When it comes to the analysis of student enrollment as a dependent variable, the estimations are conducted by the above-mentioned FEP model as well. Because zero observations are excluded from our samples, the fixed-effects OLS regression analysis

---

[42] Hausman et al. (1984) propose a fixed-effects negative binomial model for panel data by developing a fixed-effects model under full distributional assumptions of overdispersion. However, Alison and Waterman (2002) insist that because Hausman et al. (1984) does not qualify as a true fixed-effects estimator and does not maintain a good control for all unchanged independent variables, it is recommended to use the FEP model that has more robust characteristics.

[43] There is also a concern that publicly funded R&D may be correlated with the error terms through omitted variables. For example, if there are government policies that intend to strengthen the research functions of universities such as both public research funds and doctorates awarded, the estimates of R&D would be biased. However, since such policies are generally implemented via public funds, this type of the endogeneity problem is not much of a concern.

is also conducted on the basis of the robust clustered standard errors. The fixed-effects OLS model is defined as follows:

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + u_{it}. \tag{3.61}$$

The independent variables are transformed into the natural logarithms, and year dummy variables (which are omitted from the tables to save space) are also included to control for year-specific fluctuations. All universities are divided into public and private universities in order to find the inherent distinctions between the two university sectors. Several specifications are formulated for the purpose of robustness check.

Results

Tables 3.4 presents the results where the total number of doctorates awarded is a dependent variable. Both general and science R&D have positive relations with the total number of doctorates awarded only for public universities and that their coefficients are all significant. This suggests that publicly funded R&D arguably makes a considerable contribution to research output of public universities, as is normally expected from the theoretical result except that substitutability is quite strong.

By contrast, the estimates of private universities are negative and insignificant. Somewhat surprisingly, it follows that in the case of private universities, publicly funded R&D is not related to the number of doctorates awarded, or the effect might possibly be negative. One reason for this controversial result might be that as the theory suggests, producing doctorates is a powerful substitute for teaching activities, and hence, a decrease in tuition revenue undermines the budgetary foundation for research activities. This is not to be theoretically denied as demonstrated previously, but such a mechanism does not seem prevalent among overall private universities. Another more likely reason might be that the number of doctorates awarded itself is not necessarily an appropriate indicator of measuring research output. Rather, fostering doctoral students is exactly teaching activities, which could constitute a strong substitute for research activities at those private universities.

Indeed, whereas public universities are mostly homogeneous in terms of research and teaching activities, private universities are not and contain the two polar types: large prestigious universities and small educational colleges. The sample of private universities include relatively many small educational colleges that seem inclined to strong substitutability, and as a result, this kind of universities are likely to operate to affect negatively the sign of the coefficient. To validate this reasoning, we also conduct the regression analysis by eliminating observations that come from the "Ivy League" (Brown University; Columbia University; Cornell University; Dartmouth College; Harvard University; Princeton University; The University of Pennsylvania; Yale University) and other highly prestigious universities including Stanford University and Massachusetts Institute of Technology (MIT). In this analysis, the coefficient of publicly funded R&D in Estimation (9) becomes both negative (-0.154) and significant at the 10% level ($p = 0.083$). That is why private universities excluding large prestigious universities may have more of a tendency to consider research activities as being a strong substitute for teaching ones. Consequently, although publicly funded

R&D does not intend to directly increase the number of doctorates, this finding regarding private universities is a matter of importance for science and technology policies as well as university management policies, which need to provide research funds to universities aiming at more research output.

With regard to graduate tuition fee, the estimates of all universities are positive and statistically significant. So-called “cross-subsidies" may be given to doctorate research sectors from master education sectors that earn more tuition revenue than the former. [44] On the other hand, since the estimates obtained separately for public and private universities are not significant at all possibly due to the small number of observations, this argument is merely a conjecture. Finally, the estimates of the number of academic staff are not strongly significant in almost every university category (at most about 10% significance level for Estimations [1], [3], [5], [7], [9], and [11]). However, comparing the estimations between public and private universities, we find that the latter coefficients are much larger than the former. Although we have to be cautious about the effectiveness of the coefficients, it is possible that the number of academic staff (namely, a proxy of the scale of universities) is more crucial in private universities to produce doctoral students in terms of practical magnitudes.

Table 3.5 indicates the results of the number of science doctorates awarded instead of general doctorates. Like the previous estimations, the estimates of general and science R&D are positive and significant for public universities, and in particular, the coefficients of science R&D are larger and more highly significant as expected. By

[44] In general, most doctoral students receive a scholarship from their universities or other institutions such as governments or international organizations.

contrast, the estimates for private universities are all negative and almost insignificant, too. But using the same samples excluding the above-mentioned large prestigious private universities, we see that the coefficient of general R&D (-0.134) in Estimation (9) becomes significant at the 1% level ($p = 0.008$). Hence, this implies again the possibility that research and teaching activities may be strong substitutes. With regard to the number of academic staff, the estimates for private universities are weakly significant at around the 10% significance level in contrast to public universities. This also points to the importance of the university scale in creating science-related doctoral students especially in the case of private universities.

Let us move on to Table 3.6 where total student enrollment is a dependent variable and the estimations are conducted based on the FEP model. The signs of the coefficients of general and science R&D represent how public funding affects teaching activities at universities. Although the estimates for public universities are positive and those for private universities are negative, both of them are not significant. [45] The undergraduate tuition fee is negatively associated with total student enrollment for all and public universities, which implies that competition for students in general may be in force and that these universities may not be necessarily differentiated. The result of negative coefficients concerning tuition fee is consistent with existing studies (cf. Neill, 2009; Hemelt and Marcotte, 2011; Hüber, 2012). But as we will see later, the analysis of Table 3.7, in turn, reveals that the coefficients are insignificant, this result is not necessarily robust. In view of the number of academic staff, since all the esti-

[45] This study also conducts the regression analysis using the same sample excluding the large prestigious private universities as before. But although the p-values slightly improve, the estimates are still insignificant.

mates are positive and highly significant, it seems to be the major key to maintaining teaching activities.

Alarmingly, we need to notice that all the estimations do not pass the test of strict exogeneity on which the fixed-effects Poisson regression model relies, [46] and hence, we are required to prudently evaluate the above results. With this in mind, the pooled Poisson regression with robust clustered standard errors is also conducted for an auxiliary analysis. (The tables have been omitted to save space.) Although this regression model has some limitation of not taking into account the individual fixed effects of universities, such assumptions as strict exogeneity and conditional independence between a dependent variable are not required. An interesting observation is revealed that the estimates of general and science R&D (negative), undergraduate tuition fee (negative), and the number of academic staff (positive) on total student enrollment are all significant at the 5% level in every category of universities.

Table 3.6 presents the results of the fixed-effects OLS model. The dependent variable is also transformed into the natural logarithms, so that the estimation models are “log-log" forms. Interestingly enough, as for science R&D, Estimations (11) and (12) for private universities exhibit negative estimates at the 5% significance level. Meanwhile, the estimates concerning general R&D in Estimations (9) and (10) are also negative, but not significant at the 10% level. We can therefore argue that research and teaching activities of private universities may a strong substitute especially for

[46] A strict exogeneity test is performed by simply adding the one-year leads of regressors (excluding time dummies) and testing robust joint significance of the leads with an FEP estimation. A significant statistic suggests that the strict exogeneity assumption fails (Wooldridge, 2010).

science-related R&D, or that the crowding-out effect may be in force due to a capacity constraint of this university category, or both. However, we need to notice that the absolute values of the estimates are fairly small, -0.009 to -0.013. This implies that from the definition of elasticities, 1% increase in publicly funded science R&D leads to approximately only a 0.01% decrease in a total student enrollment. Furthermore, whereas neither the graduate nor undergraduate tuition fees are significant at all, the number of academic staff is still significant as a whole. (Note that only Equations [9] and [11] clearly pass the strict exogeneity test.)

Further direction for empirical research

We have briefly reviewed the empirical implications based on the theoretical discussions, but there are a few issues which attention should be paid to. The problem is that the theoretical and empirical models are somewhat far from the reality of U.S. universities, which generates some measurement and interpretation issues.

First, many U.S. universities limit student enrollment in reality; they turn away a large number of qualified applicants. One reason is that they are constrained by their capacity limitations in terms of physical and human resources, as the theory points to this possibility. Another factor which U.S. universities may have to take into account is maintaining their teaching quality – a standard quality measure is typically the head-count ratio of students to staff. If this ratio gets large, the quality of teaching will inevitably fall. Since universities usually aim to maintain their teaching quality, students may be subject to enrollment restriction even if some capacity slacks exists at universities. But in this model, the quality indicator, such as student enrollment for a fixed number of staff, is not explicitly dealt with (only teaching effort is endogenously considered). The second issue is that financial aid is very widely available for prospective students in the U.S., and most students pay only a fraction of advertised tuition fees. [47] As a result, caution should be taken when interpreting the statistical relationship between student enrollment and the tuition fee in our empirical model. Third and finally, it is noticeable that this empirical study omits from the analysis privately allocated R&D funds that are likely to have a significant impact on activities of U.S. universities.

These issues are not presently incorporated into our empirical framework partly due to unavailability of appropriate measures, but it is much more reasonable to take them into consideration for more precise analyses. The future challenge of the empirical analysis is to make the empirical specification much more realistic in order for the results to be easily interpreted.

[47] van der Klaauw (2002) demonstrates that financial aid offered by colleges significantly affect students' decisions to enroll at a particular college, and that it is important as an effective tool to compete with other colleges for students, by analyzing a sample of individuals admitted to an East Coast college in the U.S. Abraham and Clark (2006) reveal that the District of Columbia Tuition Assistance Great Program (DCTAG) provided to D.C. residents who intend to go on to higher educational institutions increased the likelihood of students applying to eligible institutions and the college enrollment rate among high school graduates. Moreover, based on *German Socio-Economic Panel*, Steiner and Wrohlich (2002) proves a significant positive effect of means-tested student aid on student enrollment in higher education.

Fixed-effects Poisson regression

	All universities				Public universities				Private universities			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: total #doctorates awarded												
log (General R&D)	−0.037	−0.030			0.081**	0.078**			−0.145	−0.124		
	(0.066)	(0.060)			(0.036)	(0.038)			(0.094)	(0.090)		
	[0.577]	[0.619]			[0.025]	[0.029]			[0.121]	[0.170]		
log (Science R&D)			−0.003	0.013			0.076**	0.075**			−0.107	−0.031
			(0.052)	(0.042)			(0.037)	(0.038)			(0.091)	(0.070)
			[0.953]	[0.746]			[0.039]	[0.048]			[0.239]	[0.659]
log (U. tuition fee)	−0.009		−0.005		0.011		0.015		−0.541		−0.059	
	(0.048)		(0.049)		(0.059)		(0.059)		(0.450)		(0.467)	
	[0.849]		[0.912]		[0.859]		[0.802]		[0.229]		[0.205]	
log (G. tuition fee)	0.089**	0.076**	0.086*	0.076**	0.045	0.049	0.041	0.048	0.057	0.052	0.043	0.036
	(0.045)	(0.036)	(0.045)	(0.035)	(0.055)	(0.038)	(0.056)	(0.038)	(0.088)	(0.083)	(0.090)	(0.084)
	[0.048]	[0.034]	[0.055]	[0.031]	[0.415]	[0.196]	[0.467]	[0.206]	[0.520]	[0.529]	[0.634]	[0.669]
log (#Acad. staff)	0.211*	0.113	0.211	0.111	0.074	0.073	0.073	0.014	0.480	0.356	0.491	0.357
	(0.128)	(0.093)	(0.130)	(0.094)	(0.049)	(0.044)	(0.049)	(0.044)	(0.312)	(0.251)	(0.325)	(0.261)
	[0.098]	[0.225]	[0.104]	[0.235]	[0.129]	[0.727]	[0.132]	[0.749]	[0.124]	[0.156]	[0.131]	[0.172]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
#Observations	2852	2947	2859	2956	1906	1939	1915	1948	946	1008	944	1008
#Universities	337	350	338	351	224	228	225	229	113	122	113	122
χ^2 statistics	304.30	300.16	301.24	277.93	286.53	299.17	286.39	300.01	60.01	56.27	55.24	55.66
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Strict exogenity test												
χ^2 statistics	4.08	2.47	3.93	2.53	2.99	2.89	2.71	2.33	6.47	6.36	5.73	5.62
	[0.395]	[0.480]	[0.415]	[0.470]	[0.560]	[0.409]	[0.601]	[0.506]	[0.167]	[0.095]	[0.220]	[0.131]

Note: 1. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

2. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.

Table 3.4. Regression analysis of the total number of doctorates awarded.

Fixed-effects Poisson regression

	All universities				Public universities				Private universities			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: #science doctorates awarded												
log (General R&D)	−0.010	−0.003			0.067*	0.066*			−0.100*	−0.080		
	(0.039)	(0.036)			(0.041)	(0.039)			(0.054)	(0.051)		
	[0.807]	[0.923]			[0.100]	[0.092]			[0.066]	[0.116]		
log (Science R&D)			0.024	0.027			0.081**	0.080**			−0.065	−0.022
			(0.037)	(0.028)			(0.040)	(0.040)			(0.060)	(0.041)
			[0.518]	[0.340]			[0.043]	[0.046]			[0.279]	[0.587]
log (U. tuition fee)	0.005		0.011		0.006		0.014		−0.018		−0.055	
	(0.540)		(0.054)		(0.064)		(0.063)		(0.317)		(0.323)	
	[0.924]		[0.838]		[0.931]		[0.820]		[0.955]		[0.865]	
log (G. tuition fee)	0.019	0.018	0.014	0.018	0.005	0.005	−0.004	0.004	−0.020	−0.013	−0.028	−0.021
	(0.044)	(0.035)	(0.044)	(0.034)	(0.061)	(0.040)	(0.056)	(0.040)	(0.065)	(0.065)	(0.065)	(0.065)
	[0.672]	[0.596]	[0.757]	[0.605]	[0.938]	[0.903]	[0.467]	[0.927]	[0.763]	[0.839]	[0.666]	[0.748]
log (#Acad. staff)	0.077*	0.028	0.076*	0.027	0.034	−0.009	0.032	−0.011	0.179	0.128	0.184*	0.126
	(0.041)	(0.036)	(0.041)	(0.035)	(0.034)	(0.031)	(0.061)	(0.032)	(0.110)	(0.084)	(0.110)	(0.084)
	[0.062]	[0.423]	[0.063]	[0.441]	[0.311]	[0.772]	[0.950]	[0.737]	[0.102]	[0.126]	[0.095]	[0.134]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
#Observations	2711	2800	2709	2800	1819	1852	1819	1852	892	948	890	948
#Universities	319	331	319	331	214	218	214	218	105	113	105	113
χ^2 statistics	612.82	631.70	617.56	633.81	485.04	479.67	489.08	488.31	191.17	199.94	191.09	200.33
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Strict exogenity test												
χ^2 statistics	2.50	0.71	2.60	2.53	2.12	3.41	2.48	3.78	14.07	16.08	13.74	14.93
	[0.644]	[0.870]	[0.627]	[0.470]	[0.713]	[0.332]	[0.649]	[0.286]	[0.007]	[0.001]	[0.008]	[0.002]

Note: 1. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

2. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.

Table 3.5. Regression analysis of the number of science doctorates awarded.

Fixed-effects Poisson regression

	All universities				Public universities				Private universities			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: total student enrollment												
log (General R&D)	−0.0004	−0.001			0.001	0.0004			−0.007	−0.008		
	(0.003)	(0.003)			(0.003)	(0.003)			(0.006)	(0.001)		
	[0.888]	[0.721]			[0.750]	[0.896]			[0.250]	[0.161]		
log (Science R&D)			−0.0002	−0.001			0.001	0.001			−0.006	−0.007
			(0.003)	(0.003)			(0.004)	(0.004)			(0.006)	(0.005)
			[0.943]	[0.790]			[0.751]	[0.872]			[0.335]	[0.220]
log (U. tuition fee)	−0.033**	−0.024*	−0.032**	−0.024*	−0.042**	−0.025*	−0.041**	−0.025*	−0.063	−0.047	−0.061	−0.045
	(0.016)	(0.014)	(0.016)	(0.013)	(0.019)	(0.014)	(0.019)	(0.014)	(0.107)	(0.098)	(0.109)	(0.100)
	[0.041]	[0.079]	[0.045]	[0.080]	[0.026]	[0.074]	[0.030]	[0.076]	[0.558]	[0.629]	[0.578]	[0.648]
log (G. tuition fee)	0.010		0.009		0.019		0.019		−0.010		−0.028	
	(0.012)		(0.012)		(0.018)		(0.018)		(0.016)		(0.065)	
	[0.426]		[0.444]		[0.283]		[0.300]		[0.519]		[0.666]	
log (#Acad. staff)	0.170***	0.174***	0.169***	0.174***	0.193***	0.195***	0.192***	0.195***	0.121***	0.131***	0.122***	0.132***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.035)	(0.035)	(0.035)	(0.035)	(0.042)	(0.042)	(0.042)	(0.042)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.004]	[0.002]	[0.003]	[0.002]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
#Observations	4950	5501	4958	5511	3254	3342	3260	3349	1696	2159	1698	2162
#Universities	638	711	639	712	410	426	411	427	228	285	228	285
χ^2 statistics	557.86	562.83	557.20	562.49	438.09	430.31	435.22	428.65	221.40	228.11	216.23	222.26
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Strict exogenity test												
χ^2 statistics	81.14	83.33	72.55	74.58	79.15	78.22	75.83	74.63	12.68	13.87	9.97	11.08
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.013]	[0.003]	[0.041]	[0.011]

Note: 1. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

2. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.

Table 3.6. Regression analysis of total student enrollment (1).

Fixed-effects OLS regression

	All universities				Public universities				Private universities			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: log (total student enrollment)												
log (General R&D)	−0.003	−0.005			−0.001	−0.002			−0.009	−0.010		
	(0.003)	(0.004)			(0.003)	(0.003)			(0.007)	(0.007)		
	[0.272]	[0.131]			[0.803]	[0.490]			[0.170]	[0.135]		
log (Science R&D)			−0.006	−0.007*			−0.001	−0.003			−0.013**	−0.013**
			(0.004)	(0.004)			(0.004)	(0.004)			(0.006)	(0.006)
			[0.123]	[0.056]			[0.835]	[0.548]			[0.038]	[0.033]
log (U. tuition fee)	−0.001	−0.003	−0.001	−0.002	−0.010	−0.009	−0.010	−0.009	0.085	0.101	0.091	0.104
	(0.015)	(0.009)	(0.015)	(0.009)	(0.015)	(0.010)	(0.015)	(0.010)	(0.126)	(0.102)	(0.127)	(0.102)
	[0.927]	[0.779]	[0.931]	[0.788]	[0.506]	[0.373]	[0.526]	[0.376]	[0.503]	[0.323]	[0.475]	[0.309]
log (G. tuition fee)	−0.004		−0.004		0.001		0.0004		−0.019		−0.017	
	(0.013)		(0.013)		(0.015)		(0.015)		(0.022)		(0.022)	
	[0.728]		[0.740]		[0.958]		[0.976]		[0.390]		[0.435]	
log (#Acad. staff)	0.129***	0.141***	0.130***	0.141***	0.098**	0.106***	0.098**	0.106***	0.170*	0.179**	0.171*	0.179**
	(0.047)	(0.045)	(0.047)	(0.045)	(0.038)	(0.040)	(0.038)	(0.040)	(0.097)	(0.085)	(0.097)	(0.085)
	[0.006]	[0.002]	[0.006]	[0.002]	[0.011]	[0.008]	[0.011]	[0.008]	[0.080]	[0.036]	[0.078]	[0.035]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
#Observations	5054	5635	5059	5642	3295	3392	3300	3398	1759	2243	1759	2244
#Universities	742	845	740	843	451	476	451	476	291	369	289	367
F statistics	38.69	42.97	38.41	42.57	37.55	40.09	37.14	39.85	7.79	9.92	7.75	9.64
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Strict exogeneity test												
F statistics	5.84	8.45	5.94	8.56	7.77	8.58	7.91	8.66	1.19	2.54	1.14	2.69
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.318]	[0.057]	[0.339]	[0.046]

Note: 1. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

2. The robust clustered standard errors and *p*-values are reported in the round and square parentheses, respectively.

Table 3.7. Regression analysis of total student enrollment (2).

References

[1] Abraham, K. G. and Clark, M. A. (2006), “Financial Aid and Students’ College Decisions: Evidence from the District of Columbia Tuition Assistance Grant Program", *Journal of Human Resources*, 41(3), pp. 578-610

[2] Aghion, P., Dewatripont, M., and Stein, J. C. (2008), “Academic Freedom, Private-Sector Focus, and the Process of Innovation", *Rand Journal of Economics*, 39(3), pp. 617-635

[3] Allison, P. D. and Waterman, R. P. (2002), “Fixed-Effects Negative Binomial Regression Models", *Sociological Methodology*, 32, pp. 247-265

[4] Azoulay, P., Waverly, D. Toby, S. (2009), “The Impact of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output", *Journal of Industrial Economics*, 57(4), pp. 637-676

[5] Banal-Estañol, A. and Macho-Stadler, I. (2010), “Scientific and Commercial Incentives in R&D: Research versus Development?", *Journal of Economics and Management Strategy*, 19(1), pp. 185-221

[6] Beath, J., Owen, R. F., Poyago-Theotoky, J., and Ulph. D. (2003), “Optimal Incentives for Income-Generation in Universities: The Rule of Thumb for the Compton Tax", *International Journal of Industrial Organization*, 21(9), pp. 1301-1322

[7] Beaudry, C. and Allaoui, S. (2012), “Impact of Public and Private Research Funding on Scientific Production: The Case of Nanotechnology", *Research Policy*, 41(9), pp. 1589-1606

[8] Caballero, R., Galache, T., Goméz, T., Molina, J., and Torrico, A. (2004), “Budgetary Allocations and Efficiency in the Human Resources Policy of a University Following Multiple Criteria", *Economics of Education Review*, 23(1), pp. 67-74

[9] Canton, E. and de Jong, F. (2005), “The Demand for Higher Education in the Netherlands, 1950-1999", *Economics of Education Review*, 24(6), pp. 651-663.

[10] David, P. A. (1993), “Positive Feedbacks and Research Productivity in Science: Reopening Another Black Box", available at:
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.159.7247

[11] De Fraja, G. (2012), “A Theoretical Analysis of Public Funding for Research", CEPR Discussion Papers, Paper No. 8442, Centre for Economic Policy Research, available at:
http://cepr.org/active/publications/discussion_papers/dp.php?dpno=8442

[12] De Fraja, G. and Iossa, E. (2002), “Competition among Universities and the Emergence of the Élite Institution", *Bulletin of Economic Research*, 54(3), pp. 275-293

[13] De Fraja, G. and Valbonesi, P. (2012), “The Design of the University System", *Journal of Public Economics*, 96(3), pp. 317-330

[14] Del Rey, E. (2001), “Teaching versus Research: A Model of State University Competition", *Journal of Urban Economics*, 49(2), pp. 356-373

[15] Demski, J. S. and Zimmerman, J. L. (2000), “On ‘Research vs. Teaching’: A Long-Term Perspective", *Accounting Horizons*, 14(3), pp. 343-352

[16] Foray, D. and Lissoni, F. (2010), “University Research and Public-Private Interaction", Chapter 6 in *Handbook of the Economics of Innovation*, Volume 1, eds. Hall, B. and Rosenberg, N., Elsevier, Amsterdam, pp. 275-314

[17] Gautier, A. and Wauthy, X. (2007), “Teaching versus Research: A Multi-Tasking Approach to Multi-Department Universities", *European Economic Review*, 51(2), pp. 273-295

[18] Greene, W. H. (2011), *Econometric Analysis*, 7th edition, Pearson Education, New York, NY, pp. 857-858.

[19] Haskins, C. H. (1957), *Rise of Universities*, 3rd revision, Cornell University Press, Ithaca, NY

[20] Hausman, J., Hall, B. H., and Griliches, Z. (1986), “Econometric Models for Count Data with Application to the Patents-R&D Relationship", *Econometrica*, 52(4), pp. 909-938

[21] Hellman, T. and Perotti, E. (2011), “The Circulation of Ideas in Firms and Markets", *Management Science*, 57(10), pp. 1813-1826

[22] Hemelt, S. W. and Marcotte, D. E. (2011), “The Impact of Tuition Increases on Enrollment at Public Colleges and Universities", *Educational Evaluation and Policy Analysis*, 33(4), pp. 435-457

[23] Henderson, R., Jaffe, A., and Trajtenberg, M. (1998), “Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988", *Review of Economics and Statistics*, 80(1), pp. 119-127

[24] Hotelling, H. (1929), “Stability in Competition", *Economic Journal*, 39(153), pp. 41-57

[25] Hübner, M. (2012), "Do Tuition Fees Affect Enrollment Behavior? Evidence from a 'Natural Experiment' in Germany", *Economics of Education Review*, 31(6), pp. 949-960

[26] Jaffe, A. (1989), "Real Effects of Academic Research", *American Economic Review*, 79(5), pp. 957-970

[27] Jensen, A. R., Thursby, J., and Thursby, M. C. (2010), "University-Industry Spillovers, Government Funding, and Industrial Consulting", NBER Working Paper Series, Paper No. 15732, National Bureau of Economic Research, available at: http://www.nber.org/papers/w15732

[28] Lacetera, N. (2009), "Different Missions and Commitment Power in R&D Organizations: Theory and Evidence on Industry-University Alliances", *Organization Science*, 20(3), pp. 565-582

[29] Lach, S. and Schankerman, M. (2008), "Incentives and Invention in Universities", *Rand Journal of Economics*, 39(2), pp. 403-433

[30] Mill, J. S. (1867), *Inaugural Address delivered to the University of St. Andrews, February 1st 1867*, Longman, Green, Reader, and Dyer, London, MDCCCLXVII, available at: https://archive.org/details/inauguraladdres00millgoog

[31] Mowery, D. C., Ziedonis, A. A. (2002), "Academic Patent Quality and Quantity before and after the Bayh-Dole Act in the United States", *Research Policy*, 31(3), pp. 399-418

[32] Neill, C. (2009), "Tuition Fees and the Demand for University Places", *Economics of Education Review*, 28(5), pp. 561-570.

[33] OECD (2014), *Main Science and Technology Indicators*, updated in July 2015, Paris, available at: http://www.oecd.org/sti/msti.htm

[34] Payne, A. A. and Siow, A. (2003), "Does Federal Research Funding Increase University Research Output?", *The B. E. Journal of Economic Analysis and Policy*, 3(1), pp. 1-24

[35] Rumberger, R. W. and Thomas, S. L. (1993), "The Economic Returns to College Major, Quality and Performance: A Multilevel Analysis of Recent Graduates", *Economics of Education Review*, 12 (1), pp. 1-19

[36] Steiner, V. and Wrohlich, K. (2012), "Financial Student Aid and Enrollment in Higher Education: New Evidence from Germany", *Scandinavian Journal of Economics*, 114(1), pp. 124-147

[37] Scotchmer, S. (2013), "Patents in the University: Priming the Pump and Crowding Out", *Journal of Industrial Economics*, 61(3), pp. 817-844

[38] van der Klaauw, W. (2002), “Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach", *International Economic Review*, 43(4), pp. 1249-1287

[39] Wooldridge, J. M. (1999), “Distribution-Free Estimation of Some Nonlinear Panel Data Models", *Journal of Econometrics*, 90(1), pp. 77-97

[40] Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, The MIT Press, Cambridge, MA, pp. 325

[41] Yonetani, Y., Ikeuchi, K., and Kuwahara, T. (2013), “An Analysis on the Relationship between Input and Output in the Production of Articles in Universities: An Approach Using Web of Science and Survey of R&D" (in Japanese), National Institute of Science and Technology Policy Discussion Papers, Paper No. 89, National Institute of Science and Technology Policy, available at:
http://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-DP89-FullJ.pdf

Chapter 4. Competition Effects on Industrial Productivity: An Analysis of Japanese Industries on the Basis of the Industry-Level Panel Data

4.1. Introduction

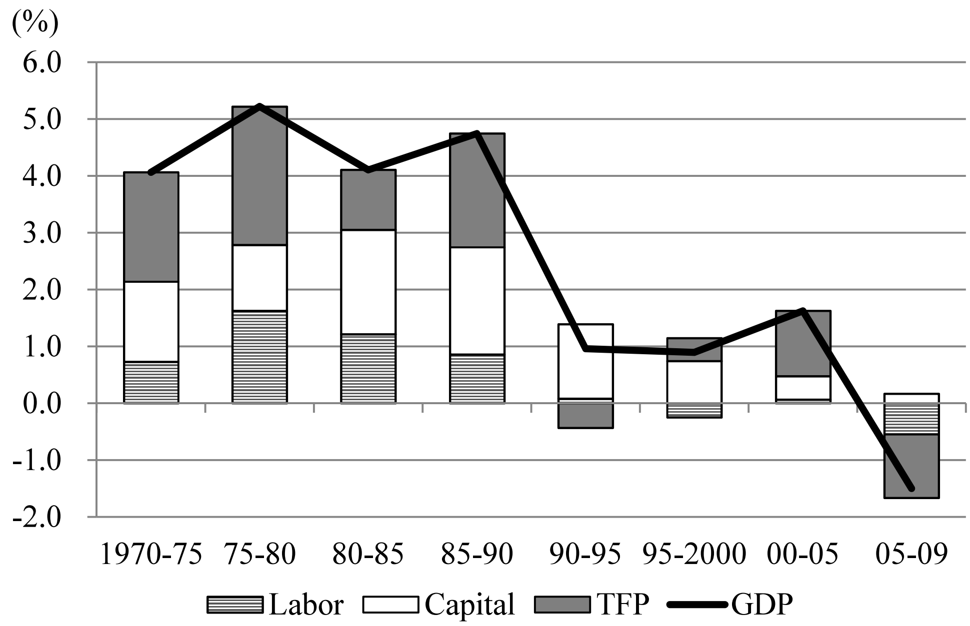
As the Structure-Conduct-Performance (SCP) paradigm notes, it has been considered that market structures such as the degree of market competition or concentration and the level of entry barriers affect the economic performance of markets. Baily and Solow (2001), who compare productivity across countries on the basis of OECD data, find large discrepancies in productivity across countries, possibly resulting from different market structures. It has been intriguing investigating which specific industrial market structures, especially the competitive environment in markets, produce the described discrepancies in productivity or growth.

With many economists addressing this complex and controversial question, two conflicting ideas regarding market competition and productivity have arisen. The first idea is that the more competitive markets generate a higher pressure to survive, that is, firms exposed to fierce market competition are forced to improve their productivity. By contrast, the second idea posits that firms in less competitive markets and having stronger market power can better afford to innovate. The complexities of market relations and characteristics necessitate empirical demonstration of the effects of market competition or market power that influence differences in productivity in a way to complement previous studies.

When attention is directed at the Japanese economy, a question arises as to why the productivity level of Japan actually continues to be extremely low. [1] As many researchers have pointed out (Hayashi and Prescott, 2002; Hoshi and Kashyap, 2011; Fukao, 2012), Figure 4.1 depicts that the average contribution of the total factor productivity (TFP) growth to real GDP growth in the Japanese economy fell sharply to below 0% between 1990 and 1995, and has remained stagnant along with production factors such as capital and labor since 1995, although it turned positive between 1995 and 2005. [2] Specifically, the TFP growth rate of the non-manufacturing industries (primarily service industries) has remained quite low for a much longer time as compared to the manufacturing industries, as illustrated in Figure 4.2. (The dataset used for Figures 4.1 and 4.2 is explained in Section 4.4.)

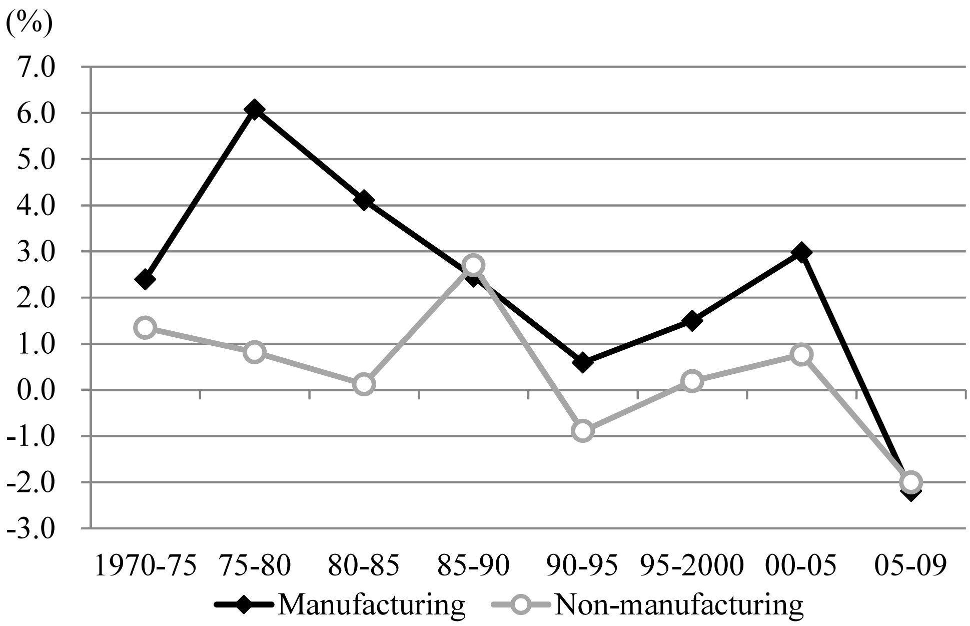
[1] According to Baily and Solow (2001), when U.S. labor productivity from 1993 to 1995 was normalized to a scale of 100, estimates for other countries were Holland, 96; West Germany, 92; France, 92; the UK, 73; Japan, 70. Furthermore, while U.S. total factor productivity (TFP) was 100 for the same period, the figures for other leading nations were West Germany, 89; France, 89; the UK, 79; Japan, 67.

[2] The conceivable reason why the average contribution of TFP growth showed an exorbitant negative figure between 2005 and 2009 is that the outbreak of the “Lehman Shock" in September of 2008 had a destructive impact on the world economies including Japan. Indeed, the Japanese TFP growth rate between 2005 and 2007 was about 1.4%.



Source: *JIP Database 2012*.

Figure 4.1. Average contribution to the real GDP growth rate.



Source: *JIP Database 2012*.

Figure 4.2. TFP growth rate of manufacturing and non-manufacturing industries.

Baily and Solow (2001), making an inter-industry comparison of productivity in manufacturing and service industries across countries, propose that although, in Japan, the export-oriented industries such as automobiles and steel exhibit high productivity, the domestic service industries have much lower productivity due to the presence of government regulations providing protection from global competition. [3] But contrasting claims have been made that the service industries in Japan, although not exposed to strong global market competition, are involved in a Bertrand-type price overcompetition in domestic markets, which hinders service industry firms from increasing their productivity. Reference is made to the causes of low productivity accruing to the non-manufacturing industries in conjunction with the estimation results in Section 4.5.

Considering the views mentioned in the preceding discussion, this chapter attempts to explore whether the idea that increased market competition improves industrial productivity is valid by analyzing their statistical relations on the basis of the Japanese industry-level panel data from 1980 to 2008. In order to develop a detailed view of the competition effect, this study breaks down the total industries into the manufacturing and non-manufacturing industries, and empirically demonstrates the difference in such effects between them. The main finding is that whereas the positive effect of market competition, calculated from the Lerner index, on TFP growth can be observed in the manufacturing industries throughout the sample period, the weak negative market competition effect may operate in the non-manufacturing industries during the latter half of the same period. This result seems to support the "Schumpeterian hypothesis" originated in Schumpeter (1942), which states that monopolistic

[3] Yoshino and Sakakibara (2002) argue that competitive strengthening of service industries is required to remedy the Japanese economy from the suffering of long-term stagnation.

firms tend to invest more heavily in R&D and create more innovations, being applied in the case of the non-manufacturing industries.

The unique contribution of this chapter to the literature on the relationship between market competition and productivity is as follows: although it employs aggregated industry-level data, and not micro firm-level data, it aims at focusing on the effect of competition on the non-manufacturing industries, which few researchers have examined so far, as well as on the manufacturing industries. The contrasting result of competition effects between manufacturing and non-manufacturing industries is somewhat intriguing and suggestive. It is critically important from the perspective of competition and innovation policies to shed light on this relationship in order to improve recent low productivity of the Japanese industries by taking into account the differences in industrial characteristics.

The remainder of the chapter is organized as follows. Section 4.2 provides a survey of existing theoretical and empirical research and reviews several empirical studies focusing on Japan. Section 4.3 defines the empirical formulations while describing an endogeneity problem. Section 4.4 explains the construction of the variables. Section 4.5 reports empirical results and their interpretations in reference to other studies. Section 4.6 concludes and discusses significant implications. Section 4.7 lists appendices followed by full references.

4.2. Survey of existing studies

4.2.1. Theoretical backgrounds

In the case where technology is assumed to be appropriated, a simple reasoning suggested by Arrow (1962) points out that firms in a competitive market generally have stronger incentives to achieve technological progress that reduces costs than monopoly firms. More precisely, competitive firms are eager to innovate in order to achieve the status enjoyed by monopoly firms and to earn monopoly profits by owning a breakthrough innovative technology. By contrast, monopoly firms remain in unchanged market positions even after achieving their own technological progress, and hence, the incentive to further innovate would weaken. This mechanism in monopoly firms is often called the “replacement effect". [4]

By contrast to the “static" efficiency of perfect competitive markets, Schumpeter (1942) highlights the importance of the “dynamic" problem. As Ahu (2002) summarizes, Schumpeter’s (1942) argument is that the organization of firms and markets that is most conducive to solving the static problem of resource allocation is not necessarily most conducive to rapid technological progress. Hence, Schumpeter (1942) concludes that firms typically operating in competitive markets are not as dynamically efficient as large firms operating in more concentrated markets. [5] His work

[4] Arrow (1962) also notes the possibility that if technology is not appropriable, the amount of R&D expended for technology is less than the socially optimal level because firms want a free ride on the R&D outcomes achieved by other firms without facing the burden of expenses. See also the discussion developed in Section 1.1 of Chapter 1.

[5] Cohen (1995) provides numerous reasons for the advantage enjoyed by large firms in concentrated markets that engage in R&D, illustrating the examples of capital market imperfections, fixed costs of innovation (particularly process innovation), complementarities between R&D activity and non-manufacturing activity, and diversification permitting economies of scale or risk reduction. Con-

is reinterpreted as the "Schumpeterian hypothesis" by later economists who consider monopoly power conducive to the progress of innovative activity. Inspired by the intuitive works of Schumpeter (1942) and others, many economists have conducted theoretical and empirical studies to test whether or not the Schumpeterian hypothesis holds true, particularly, in terms of whether competition (or monopoly) promotes growth, technological progress, and innovation.

In addition to early theoretical works (Dasgupta and Stiglitz, 1980; Gilbert and Newbery, 1982), the contract theory approach is widely used to assess the relationship between competition and productivity. Hart (1983) reveals that if high-incentive entrepreneurial firms initiate a general reduction in costs and prices, low-incentive managerial firms are also compelled to engage in cost cuts with a reduction in managerial slack as they are confronted with the threat of more efficient rivals. Thus, Hart (1983) suggests that competition in a product market reduces managerial slack and improves productivity. By contrast, Scharfestein (1988) argues that market competition, with imperfect information, between entrepreneurial (the productivity of a manager is observable) and managerial firms (it is non-observable) may instead exacerbate incentive problems. The entrepreneurial firm can decrease its price (increase output) by becoming more efficient when its productivity is low, and accordingly, the target profit of the managerial firm set by the owner is lowered across the board. But this gives a high-productive managerial firm the incentive to pretend that its productivity is low. As a result, competition with entrepreneurial firms can cause a negative external-

trastively, Scheler and Ross (1990) point to the counter-argument, identifying disadvantages such as excessive bureaucratic control and scientists' or entrepreneurs' low morale in large firms.

ities to the efficiency of managerial firms. Additionally, Schmidt (1997) points to the trade-off problem of competition by demonstrating that whereas increased competition reduces the profits of firms and forces managers to work harder toward improving productivity to avoid liquidation, a reduction in profits also deteriorates the profitability of cost-reduction. In short, such theoretical analyses are overall inconclusive as to the simple effect of market competition on productivity and innovative activities, depending on the researchers' assumptions and model frameworks.

Meanwhile, Aghion et al. (2005) theoretically prove that the relationship between aggregate innovation and the degree of competition can take an inverted-U shape. These studies insist that the inverted-U shape results from a combination of both the “escape-competition effect" and the “Schumpeterian effect" among heterogeneous firms. More precisely, the former effect indicates that more competition motivates firms in neck-and-neck sectors to innovate in order to escape the competition, and the latter indicates that an increase in competition discourages firms in unlevel sectors to innovate because of the dissipation of rents that can be captured by a follower after innovation. Hence, this theory can be interpreted as partially incorporating the Schumpeterian hypothesis into the model, which suggests a positive relation between market power and innovation.

4.2.2. Empirical studies

Let us turn our attention to existing empirical studies. Most notably, Nickell (1996) investigates how market environments, for example, market share, market concentration, rent (the Lerner index or price-cost margin), and the number of competitors, affect the TFP level and TFP growth by estimating a production function including these independent variables from the data of roughly 700 UK manufacturing firms between 1972 and 1986. This study reveals that market power, represented by the market share, reduces the TFP level and that market competition, represented by the Lerner index, is associated with higher rates of TFP growth. Geroski (1990), conducting a regression analysis on the UK data of 73 industrial sectors from 1970 to 1979, shows that a rise in market concentration reduces the number of innovations, and hence, concludes that there is little support for the Schumpeterian hypothesis. Blundell et al. (1999), who designate counts of innovation and patents as dependent variables from the data of 340 UK manufacturing firms gathered between 1972 and 1982, find that increased product market competition in the industry measured by market concentration tends to stimulate innovative activities, although market share has a robust positive effect on headcounts of innovations and patents. Contrastively, Crépon et al. (1998), using the cross-sectional data of innovation output of French manufacturing industries in 1990, demonstrate that the probability of conducting R&D increases significantly with firm size, market share, and diversification, as suggested by the Schumpeterian hypothesis.

Furthermore, Aghion et al. (2005) test the theoretical result of an inverted-U relation between market competition and innovation based on a panel dataset of 311 UK firms from 1973 to 1994. While citation-weighted patent count is used as a dependent variable, competition index calculated by the Lerner index and the square of this index are used as independent variables. Constructing industry-specific variables from these datasets, their study shows that the coefficient of the squared competition index is significantly negative and that the upward-sloping segment of an inverted-U shape is steeper if the set of industries is restricted to those falling below the median technological gap (all the more so in the case of “neck-and-neck" industries). [6] In a few words, their latter finding illustrates that competition effects on innovation are intensified in close proximity to technological frontiers.

A limited number of studies on Japan have been conducted centering on the relationship between competition and productivity, largely because of a lengthy delay in the establishment of a reliable database. Nevertheless, prominent research has appeared in recent years mainly using the firm-level data. Okada (2005), following Nickell’s (1996) empirical approach and using the *Basic Survey of Business Structure and Activities* (*BSBSA*) data of roughly 100,000 manufacturing firms from 1994 to 2000, demonstrates that competition measured by the lower Lerner index at the industry level reinforces productivity growth and that market power measured by either the Lerner index or market share at the firm level negatively affects the productivity level of firms performing R&D. Focusing on both productivity and innovative activity using the *BSBSA* data of about 2,400 firms from 1994 to 2001, Motohashi et al. (2005) reveal that a drop in the Herfinahl index has a positive impact on productivity but a negative impact on R&D expenditure and the number of registered patents.

Arai (2005), who uses the *Japan Industrial Productivity Database* (*JIP Database*) of 84 industrial sectors from 1970 to 1998, maintains that many sectors exhibit a pos-

[6] Scherer (1965) produces an initial empirical study that finds a non-linear relation between market structures (i.e. firm size and concentration ratio) and innovative outputs (i.e. patents).

itive correlation between the TFP growth rate and the approximated Lerner index, while the inverted-U relation is statistically observed in very limited sectors. Because of this, Arai (2005) concludes that competition may not have a positive effect on productivity. Flath (2011) uses the industry-level data produced by the *Census of Manufacturers* from 1961 to 1990, and demonstrates that there is a “U-shape" relation between market concentration measured by the Herfindahl index and technological growth (the horizontal axis is measured by the “negative" degree of competition), but this has no relation to the Lerner index. Inui et al. (2012), based on firm-level data of roughly 35,000 observations between 1997 and 2003 produced by the *BSBSA*, investigates whether the inverted-U shape theory applies in the case of Japanese manufacturing firms following the study of Aghion et al. (2005). In controlling for an endogeneity problem, this study shows not only that market competition measured by the Lerner index positively affects productivity growth, but also that there exists an inverted-U relation between them. Finally, Yagi and Managi (2013) also empirically find the inverted-U shape adopting patent data as a dependent variable and price-cost margin as an independent variable on the basis of firm-level and industry-average data from 1964 to 2006.

Economists have yet to reach an overwhelming consensus, including studies conducted in Japan, due to the difficulty in choosing the appropriate measurement variable for competition and identifying the causal relation. However, it seems that recent studies have found market competition to have a positive effect on productivity and innovative activity in manufacturing industries, resulting in disproportionate evidence against the Schumpeterian hypothesis.

4.3. Empirical formulation

This chapter's empirical formulation of the relationship between competition and productivity is close to, in particular, Okada (2005) and Inui et al. (2012). This study uses an industry-specific competition measure and TFP growth, and exploits the industry-level *JIP Database*. (A detailed explanation on the dataset is given in Section 4.4.) It should be noted that our empirical formulation differs from previous studies, in that it adds not only an index that indicates the degree of market competition (as measured by an approximated industry-level Lerner index), but also other control variables such as the incremental research and development (R&D) stock to output ratio and the IT investment to output ratio that can directly affect the industrial productivity. [7] Furthermore, by adding the quadratic term of the competition measure as other researchers do, this study intends to test the idea proposed by Aghion et al. (2005) that the competition-innovation relation takes an inverted-U shape.

All industries are split into the manufacturing and non-manufacturing industries as described later, so that the focus can be directed at specific characteristics prevalent in each industrial category. In other words, this chapter attempts to investigate how the degree of competition and productivity-related factors affect TFP growth for both

[7] Aside from these variables, indicators of trade openness, such as import penetration and foreign direct investments (FDIs), can affect productivity. But since the objective of this paper is to investigate the relation between the competition measure and the TFP growth rate, our analysis is concentrated on market competition and the other directly (more plausibly) relevant independent variables, that is, R&D and IT investment.

manufacturing and non-manufacturing industries in Japan. In particular, the focus on the non-manufacturing industries is the contribution of this study to the literature.

First, in order to simply test whether the effect of increased competition is positive or negative, the basic regression model is defined as follows:

$$tfpg_{it} = \alpha_i + \alpha_t + \beta_1 comp_{it-1} + \beta_2 \Delta rds_{it-2} + \beta_3 tit_{it-2} + \varepsilon_{it}$$
$$\text{for } i = 1, \ldots, I \text{ and } t = 3, \ldots, T. \tag{4.1}$$

$tfpg$ is the annual TFP growth rate, $comp$ is the degree of competition, Δrds is the incremental R&D stock to output ratio, tit is the total IT investment to output ratio, i is the industry script, t is the time script, α_i is the industry fixed effects, α_t is the time fixed effects, vector β_j ($j = 1, 2$, and 3) denotes the population regression coefficients, and ε_{it} is the serially uncorrelated random error terms. The variables used in this analysis are briefly summarized in Table 4.1 below.

Variables	Definition
Dependent variable	
$tfpg$	Total factor productivity (TFP) annual growth rate (%)
Independent variables	
$comp$	Degree of competition (%) calculated by 1 − *Lerner index* (%) Lerner index is calculated by 1−(*intermediate input*+*labor input*+*capital service input*)/*output* (all values are evaluated by nominal prices)
$comp^2$	Square of $comp$
rds	Ratio of nominal research and development (R&D) stock to nominal output (%)
tit	Ratio of nominal total IT investment to nominal output (%)
Instrumental variables	
$hcons$	Ratio of household consumption to nominal output (%) (obtained from final demand by sectors)
exp	Ratio of export to nominal output (%) (obtained from final demand by sectors)

Source: *JIP Database 2012* and *Estimation of Industry-Level R&D Stock*.

Table 4.1. Summary of variables.

It is posited that the one-year lag of the competitive measure and the two-year lag of incremental R&D stock and IT investment affect present-time TFP growth. This premise of the one-year lag of the competition measure is the same as that posited by Inui et al. (2012). Although many other studies assume that TFP growth and the degree of competition are concurrently related, it seems more plausible that the effect of competition would come in force in due time, especially in the case of industrial analyses. Incremental R&D stock and total IT investment are also assumed to take a prolonged period of time to have any influence on productivity. The two-year lag of these two control variables are comprehensively determined by investigating the correlation between these variables and TFP growth. [8] [9]

If competition is assumed to stimulate industries in improving productivity, β_1 will be positive. Inversely, β_1 being negative suggests that increasing market power may stimulate productivity improvement, which lends support for the Schumpeterian hypothesis. With regard to the coefficients of the incremental R&D stock to output ratio and the total IT investment to output ratio, it is generally expected that β_3 and β_4 are positive.

The following model that adds the quadratic term of the competition measure, $comp^2_{it-2}$, is also estimated to test the inverted-U shape theory:

$$tfpg_{it} = \alpha_i + \alpha_t + \beta_{11} comp_{it-1} + \beta_{12} comp^2_{it-1} + \beta_2 \Delta rds_{it-2} + \beta_3 tit_{it-2} + \varepsilon_{it}. \tag{4.2}$$

Regarding the signs of the coefficients, β_{11} and β_{12} are expected to be positive and negative, respectively, according to this theory.

Fixed-effects (FE), or within-group transformation of Equations (4.1) and (4.2) can be made to eliminate the industry fixed effects. By this transformation, for example,

---

[8] The correlations between the concurrent, one-year lag, and two-year lag of the incremental R&D stock to output ratio and the concurrent TFP growth rate are calculated as follows: −0.141 (p = 0.000); 0.121 (p = 0.000); and 0.195 (p = 0.000), respectively. With regard to the total IT investment to output ratio, the relevant correlations are calculated as: 0.007 (p = 0.742); 0.031 (p = 0.128); and 0.045 (p = 0.032), respectively. Although it is possible to assume that the further lags of the control variables can be included into Equation (4.1), this has a drawback of decreasing observations that are used in estimations. For this reason, we do not include the further lags taking into consideration the limited number of observations in our dataset.

[9] See also the later discussion that argues the advantage of taking a lag of the incremental R&D stock to output ratio and the total IT investment to output ratio from the viewpoint of the endogeneity problem.

Equation (4.1) is modified as follows:

$$tfpg_{it} - \overline{tfpg_i} = (\alpha_t - \overline{\alpha}) + \beta_1(comp_{it-1} - \overline{comp_i}) + \beta_2(\Delta rds_{it-2} - \overline{\Delta rds_i})$$
$$+ \beta_3(tit_{it-2} - \overline{tit_i}) + (\varepsilon_{it} - \overline{\varepsilon}_i), \tag{4.3}$$

where the “bar" notations denote the operation of taking the mean over time. This formulation is an example of a classical regression model that removes unobservable individual fixed effects. One advantageous feature of FE transformation is that, as long as the independent variables are uncorrelated with the error terms, ε_{it}, we can obtain consistent FE estimators even when the independent variables are correlated with the industry fixed effects. [10]

We must, though, consider that an endogeneity problem can occur when we intend to run such a regression as the above equations. [11] It may be problematic estimating Equations (4.1) and (4.2) based on the simple FE model without using instrumental variables (IVs), because the degree of competition is likely to be correlated with the error term. In particular, reverse causalities, which trigger a simultaneity bias, seem to exist between the degree of competition and the annual TFP growth rate. If independent variables are correlated with the error term, estimators are generally biased

[10] Although the alternative way of estimating the model with individual unobservable effects is to take a first difference (FD), there seems little reason to prefer one over the other. But FE (within) estimators are usually favored in a static model as they are more efficient if ε_{it} is not serially correlated. In addition, FE estimators are considered having an advantage over FD estimators for a large time dimension when contemporaneous exogeneity holds but strict exogeneity fails (Wooldridge, 2010). Subsection 4.7.3 briefly discusses the results of FD estimation.

[11] Many recent studies have supported the viewpoint that competition (or market power) and innovative activity are simultaneously determined (Cohen, 2010). For instance, Symeonidis (1996), who carefully surveys research conducted on this relation, summarizes: market structures and R&D intensity are jointly determined by technology, demand characteristics, the institutional framework, strategic interaction and chance.

and inconsistent. As Nickell (1996) and Okada (2005) stress, the reverse causality between competition and productivity is expected to generate the opposite sign. That is, the effect of productivity on competition (or market power) is likely to be negative (or positive). According to Okada (2005), if a positive relation between competition and productivity (or a negative relation between market power and productivity) is observed, competition would have a much stronger effect on productivity. It is assumed in this study that the degree of competition is predetermined for one year before TFP grows. However, if this competition measure is serially correlated, then the one-year lag of the competition measure would be also correlated with the error term, as Inui et al. (2012) point out. [12] Since this may generate an endogeneity problem, we should employ exogenous IVs from the model to circumvent it. [13]

As regards to incremental R&D stock and total IT investment, we cannot completely deny the possibility of an endogeneity problem accruing to these control variables, either. More precisely, TFP growth might be higher for those industries that conduct R&D and IT investment because they can afford to engage in more such activities due to their high productivity. Unfortunately, however, it is difficult to obtain in our dataset the valid IVs that allow us to appropriately estimate the model. Nevertheless, the degree of endogeneity generated by these variables does not seem quite as serious as the competition measure for the reason that they are included into the

[12] Moreover, the strict exogeneity assumption, $\mathbb{E}(\varepsilon_{it}|comp_{i1}, comp_{i2}, \ldots, comp_{iT-1})$ for $t = 1, 2, \ldots T$, may not be guaranteed either.

[13] Aghion et al. (2005) find the policy instruments represented by the introduction of policy changes that generated exogenous variation in the degree of industry-wide competition. Instead of using such policy instruments, Nickell (1996) and Okada (2005) estimate their models based on the Arellano-Bond generalized method of moments (GMM) estimation (Arellano and Bond, 1991) in the form of a dynamic panel data model.

estimation model by accounting for a two-year lag, so long as the serial correlations of these variables are not strong. [14]

The following variables are employed as the IVs of the competition measure in $t-1$ ($comp_{it-1}$): the change of the competition measure from $t-2$ to $t-1$ ($\Delta comp_{it-1}$) and from $t-3$ to $t-2$ ($\Delta comp_{it-2}$), the ratio of household consumption to output in $t-2$ ($hcons_{it-2}$), and the ratio of export to output in $t-2$ (exp_{it-2}), all of which are calculated at the industry level. Note that different IVs are employed in Equations (4.1) and (4.2) as described in the notes of Tables 4.5 to 4.10. Here it is considered that these IVs affect only the competition measure, but not error terms. In particular, prior final demand of household consumption and exports relative to output within industries seem to represent the market structures that can be related to the degree of competition. [15] In order to simply confirm whether the IVs are usable, the correlation coefficients between the independent variables and these potential IVs are calculated, as presented in Table 4.2. The result demonstrates that these IVs are largely correlated with the competition measure. Furthermore, the exogeneity, underidentification, weak identification, and overidentification tests are conducted in estimating the model to check the adequacy for conducting the FE-IV estimation.

[14] Wei and Liu (2006) assert that one way of keeping the possible endogeneity problem that exists between productivity and R&D (and other variables such as exports) to a minimum extent is to take a lag of the R&D variable, which would affect productivity with a time lag. The authors estimate the effect of R&D, exports, and variables regarding knowledge spillovers from exports and FDIs on TFP growth by using a one-year lag of these independent variables.

[15] A note of consideration is that there is some debate on using the ratio of export to output as an IV for the competition measure because export experience could directly contribute to TFP growth (i.e. learning by exporting). In addition, although existing studies such as Inui et al. (2012) employ the import penetration ratio as an IV, this study cannot construct a valid IV of this index obtained from the dataset.

	$comp_{t-1}$	Δrds_{t-2}	tit_{t-2}
$\Delta comp_{t-1}$	0.059***		
	[0.005]		
$\Delta comp_{t-2}$	0.074***	0.132***	0.003
	[0.000]	[0.000]	[0.898]
$hcons_{t-2}$	−0.177***	−0.067***	−0.047**
	[0.000]	[0.002]	[0.024]
exp_{t-2}	0.037*	0.180***	−0.045**
	[0.075]	[0.000]	[0.029]

Note: 1. These correlation coefficients are calculated for the sample period between 1980 and 2008.
2. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The p-values are reported in the square parentheses.

Table 4.2. Correlation between independent variables and IVs.

4.4. Dataset

This section provides detailed explanations of how the variables are constructed. Similar to Arai (2005), this study uses the *JIP Database* to obtain the primary indicators. The *JIP Database* is produced jointly by the Research Institute of Economy, Trade and Industry (RIETI) and Hitotsubashi University for the purpose of studying changes in the industrial structure of the Japanese economy. The *JIP Database* is comprised of various types of annual datasets that are necessary for estimating sectoral TFP in 108 industries covering the Japanese economy as a whole. These datasets include output and input (nominal and real levels), capital and labor costs, capital and labor input indices (labor input is adjusted taking labor quality into account), and the normalized TFP level. The *JIP Database 2012* is the primary source for the collecting of data on the industry-specific TFP growth rate, output, intermediate, labor, and capital input costs, and final demand such as consumption and exports. [16] The advantage of using the *JIP Database* is that, as Arai (2005) notes, detailed long-term data of the Japanese industrial sectors enables us to investigate the change and trend of production, productivity (TFP), and other important factors in both manufacturing and non-manufacturing industries.

The object of the analysis in this study is the 86 industrial sectors (54 manufacturing industries and 32 non-manufacturing industries) listed in Table 4.3. Not only the 14 industrial sectors that are not based on the market economy, such as social insurance/welfare, education, and medical, but also the 6 industrial sectors related to the primary industries such as agriculture, forestry, and fishing, are excluded from the sample. This is because these non-market and primary sectors are not sufficiently exposed to market competition and are frequently protected by regulations. Two industrial sectors, housing and the unclassified sectors, are also excluded due to constraints of data availability. The empirical analysis is conducted in accordance with all, manufacturing, and non-manufacturing industries in order to capture their individual characteristics.

While Arai (2005) analyzes each industrial sector from 1970 to 1998 not dividing the sample period, the period subject to estimation in our study is the years from 1980 to 2008 (inclusive). Since our analysis is subject to this long-term period, the entire period is divided into the two categorical periods: 1980-1994 and 1995-2008. As indicated in Figures 4.1 and 4.2, the Japanese economy experienced a drastic decline

[16] This brief explanation of the *JIP Database* is based on the homepage of the RIETI website. See the following page for details available at: http://www.rieti.go.jp/en/database/JIP2012/index.html.

in GDP and TFP growth since the 1990s caused by the bubble economy burst. Indeed, many economists have reached an agreement that there were some structural changes within the Japanese industries around this time, such as more competitive economic environments at a global level. This is why it is much more meaningful to examine how the difference in productivity was generated by competition before and after this period. [17]

[17] The way of splitting the time period of panel data may be subject to debate. Nevertheless, there seem to be some plausible reasons to conduct analyses dividing the dataset into 1980-1994 and 1995-2008. First, as was shown in Equation (4.1), since the estimation model includes a two-year lag of the differenced variable (i.e. Δrds_{t-2}), the period subject to estimation of TFP growth in the latter half period substantially starts from 1998. When investigating the real GDP growth rate (SNA database) and TFP growth rate (*JIP Database* excluding the housing and unclassified sectors) in the 1990s, we see that these two figures fell sharply by -1.13% and -1.03% in 1998, respectively, because of shocks delivered by the economic crisis. Moreover, we can also find that the degree of competition changed largely between 1980-1994 and 1995-2008 as described later in Subsection 4.4.3. Second, the estimation result regarding the negative competition effect on the non-manufacturing industries, which contributes a main focus of this paper, is the most conspicuous if we use the dataset of 1995-2008 (Subsection 4.5.3). Third, the test for poolability (Chow test) is conducted in reference to Baltagi (2008), althoug it is not much rigorous due to the inclusion of lagged variables. The test F-statistics exhibits $F = 10.978$, which is distributed with the degree of freedom $(340, 1892)$, so that the null hypothesis of poolability is rejected at the 1% significance level. Hence, we conclude that this panel data is not poolable and that dividing it in the year between 1994 and 1995 can be accepted.

Index	M/NM	Industry name
7	NM	Mining
8	M	Livestock products
9	M	Seafood products
10	M	Flour and grain mill products
11	M	Miscellaneous foods and related products
12	M	Prepared animal foods and organic fertilizers
13	M	Beverages
14	M	Tobacco
15	M	Textile products
16	M	Lumber and wood products
17	M	Furniture and fixtures
18	M	Pulp, paper, and coated and glazed paper
19	M	Paper products
20	M	Printing, plate making for printing and bookbinding
21	M	Leather and leather products
22	M	Rubber products
23	M	Chemical fertilizers
24	M	Basic inorganic chemicals
25	M	Basic organic chemicals
26	M	Organic chemicals
27	M	Chemical fibers
28	M	Miscellaneous chemical products
29	M	Pharmaceutical products
30	M	Petroleum products
31	M	Coal products
32	M	Glass and its products
33	M	Cement and its products
34	M	Pottery
35	M	Miscellaneous ceramic, stone and clay products
36	M	Pig iron and crude steel
37	M	Miscellaneous iron and steel
38	M	Smelting and refining of non-ferrous metals
39	M	Non-ferrous metal products
40	M	Fabricated constructional and architectural metal products
41	M	Miscellaneous fabricated metal products
42	M	General industry machinery
43	M	Special industry machinery
44	M	Miscellaneous machinery
45	M	Office and service industry machines
46	M	Electrical generating, transmission, distribution and industrial apparatus
47	M	Household electric appliances
48	M	Electronic data processing machines, digital and analog computer equipment and accessories
49	M	Communication equipment
50	M	Electronic equipment and electric measuring instruments

Table 4.3. Industry list.

Index	M/NM	Industry name
51	M	Semiconductor devices and integrated circuits
52	M	Electronic parts
53	M	Miscellaneous electrical machinery equipment
54	M	Motor vehicles
55	M	Motor vehicle parts and accessories
56	M	Other transportation equipment
57	M	Precision machinery and equipment
58	M	Plastic products
59	M	Miscellaneous manufacturing industries
60	NM	Construction
61	NM	Civil engineering
62	NM	Electricity
63	NM	Gas, heat supply
64	NM	Waterworks
65	NM	Water supply for industrial use
66	NM	Waste disposal
67	NM	Wholesale
68	NM	Retail
69	NM	Finance
70	NM	Insurance
71	NM	Real estate
73	NM	Railway
74	NM	Road transportation
75	NM	Water transportation
76	NM	Air transportation
77	NM	Other transportation and packing
78	NM	Telegraph and telephone
79	NM	Mail
81	NM	Research (private)
85	NM	Advertising
86	NM	Rental of office equipment and goods
87	NM	Automobile maintenance services
88	NM	Other services for businesses
89	NM	Entertainment
90	NM	Broadcasting
91	NM	Information services and internet-based services
92	NM	Publishing
93	NM	Video picture, sound information, character information production and distribution
94	NM	Eating and drinking places
95	NM	Accommodation
96	NM	Laundry, beauty and bath services
97	NM	Other services for individuals

Note: 1. Index corresponds to that of *JIP Database 2012*.
2. M and NM denote the manufacturing and non-manufacturing industries, respectively.

Table 4.3. Industry list (continued).

4.4.1. Dependent variable (TFP growth)

The TFP growth rate is employed as a dependent variable that encompasses innovation. The data of industry-level TFP is readily available from the *JIP Database 2012*. [18] In fact, R&D expenditure or intensity was widely used as a measure of innovative activities in earlier studies, such as Cohen and Klepper (1996), which found a positive correlation between R&D activities and firm size. And yet, the use of such measures as productivity and innovation counts, which allow us to directly comprehend the result of innovative activities, has become the preferred method recently (Cohen, 2010). For the purpose of robustness check, a preliminary regression analysis is conducted, where the real R&D investment growth rate is a dependent variable and the one-year lag of both the competition measure and the R&D stock to output ratio are independent variables. (The details are omitted to save the space.) But significant evidence of competition effects cannot be obtained from this regression as found in the next section.

On the other hand, there is some question as to whether industry-level TFP is an appropriate indicator for testing the Schumpeterian hypothesis regarding innovation. Indeed, while industry-level TFP growth utilized in this study is decomposed into productivity dynamics comprising of the internal, distribution, entry, and exit effects, the conventional Schumpeterian hypothesis generally views only the internal effect (that is, productivity improvement inside firms) as a result of innovative activities. However, since the internal effect accounts for a large part of sectoral TFP growth, we

[18] See the RIETI homepage listed in Footnote 16 or Fukao and Miyagawa (2008) for details on how to estimate industry-specific TFP.

can regard TFP growth as an approximate measure of innovation. [19] Or it may well be that we define the industry version of the Schumpeterian hypothesis as including all productivity dynamics that reflect industrial refreshment.

4.4.2. Independent variables and IVs

With regard to independent variables, the major indicator of competition is considered to be the Lerner index (the price-cost margin) used by numerous researchers (Nickell, 1996; Okada, 2005; Aghion et al., 2005; Arai, 2005; Flath, 2011; Inui et al., 2012). Indeed, as Arai (2005) points out, the conventionally popular measures of market competition in the context of competition policy are the Herfindahl index and concentration ratio. But they can appropriately reflect only the Cournot-type quantity competition where an increase in firms intensifies competition and not the Bertrand-type price competition where a decrease in firms coexists with fierce competition. In conducting a regression analysis, a practical predicament that the Herfindahl index and concentration ratio are available only for three years (1996, 2001 and 2006) in the *JIP Database 2012* prevents us from accumulating a sufficient number of observations. Although a preliminary regression analysis reveals that the Herfindahl index as an independent variable has positive correlation with the concurrent TFP growth rate (that is, high TFP growth may induce a larger market share). In common with the above-mentioned existing studies, this chapter, therefore, employs the Lerner index as a measure of competition calculated for each industrial sector.

[19] Using the financial data of Japanese listed companies, Kim et al. (2010) make it clear that an increase in TFP consistently resulted mainly from the internal effect since the 1980s including both manufacturing and non-manufacturing industries and that the other effects were relatively minute.

According to the basic definition, the Lerner index is defined as $(p-MC)/p$, where p is the price and MC is the marginal cost, and hence, this index measures a certain type of monopoly rent or profitability that implies some market power. Because it is difficult to directly calculate the marginal Lerner index based on this definition, we define the following industry-specific Lerner index as Arai (2005) does:

$$LI = \frac{output - intermediate\ input - labor\ input - capital\ service\ input}{output}, \qquad (4.4)$$

where all variables are evaluated by nominal prices and all data is available from the *JIP Database 2012*. See Subsection 4.7.1 for the background of monopoly rent and the industry-specific Lerner index. This Lerner Index for manufacturing and non-manufacturing industries, which reflects industrial profitability, adequately captures differences in both global and domestic competitive environments.

Based upon the above construction of LI, the industry-level degree of competition can be simply defined as $comp = (1 - LI) \times 100$ (%), which means that the larger the value, the more competitive the relevant industry. Since it is highly likely that the Lerner index (competition measure) fluctuates with business cycle either pro-cyclically or counter-cyclically, [20] year dummy variables are included as independent variables to control demand fluctuations.

The basic R&D data has been obtained from the *Estimation of the Industry-Level R&D Stock* edited by the National Institute of Science and Technology Policy (NIS-

[20] Green and Porter (1984) demonstrate that the Lerner index moves in accordance with business fluctuation, that is, it rises in economic booms and declines in recession (i.e. pro-cyclical). By contrast, Rotemberg and Saloner (1986) predict that the Lerner index moves in the opposing direction of business fluctuation (i.e. counter-cyclical).

TEP). This estimated data accumulates long-run deflated R&D stock (namely, technological knowledge stock) at the industry level ranging from 1973 to 2008, and the classification of industrial sectors is adjusted to be the same as those of the *JIP Database*. Considering the fact that flow of R&D affects TFP growth, this study employs as an independent variable incremental R&D stock normalized by nominal output. Flow data regarding IT investment (including both hardware and software investment) in each industrial sector between 1970 and 2008 is also provided by the *JIP Database 2012*. Finally, with regard to the consumption and export to output ratios at the industry level, the data can be obtained from final demand by sectors in the inter-industry relations table included in the *JIP Database 2012*.

4.4.3. Descriptive statistics

Descriptive statistics of these variables such as the mean and standard deviation are provided in Table 4.4, where industries are categorized into all, manufacturing and non-manufacturing industries for each period. Although the degree of competition must theoretically take values ranging from 0 to 100 in percent figures, it actually takes values beyond 100 in the non-manufacturing industries due to the negative values of the Lerner index. There are some reasons for this (Fukao et al., 2011). First, since total output and intermediate, labor and capital service inputs are separately estimated from the micro data and the estimation is not modified on the whole, the numerator of Equation (4.4) (i.e. monopoly rent) may take negative values. Second, if firms hold excess labor and capital and they are slow to adjust, then there is a tendency for the estimates of labor and capital service input to have upward bias, and whereby, the

Lerner index can be consequently negative. For these reasons, we simply use a negative value of the Lerner index and the resultant degree of competition for estimations, instead of arbitrarily transforming negative values of the Lerner index to zero.

Table 4.4 reveals that the time mean over the industries of almost every variable, especially the TFP growth rate and the degree of competition, differ statistically between the manufacturing and non-manufacturing industries. Whereas the TFP growth rate achieved by the manufacturing industries is much higher than that of the non-manufacturing industries, the degree of competition in the manufacturing industries is lower than the non-manufacturing industries. This outcome seems to reflect the fact that while the manufacturing industries have achieved a steady improvement in productivity along with satisfactory profits, the non-manufacturing industries have suffered low productivity and weak profits for a lengthy period of time. We can also find that the extent to which market environments got competitive for both the manufacturing and non-manufacturing industries is observably larger in the period of 1995-2008 than that of 1980-1994, as the change in the degree of competition indicates. [21]

[21] As shown in Table 4.4, the differences in the degree of competition with regard to all, manufacturing, and non-manufacturing industries in 1980-1994 (1995-2008) are −0.220% (0.405%), 0.039% (0.349%), and −0.615% (0.491%), respectively.

	All	Manufacturing	Non-manufacturing	*t*-value	*p*-value
Year: 1980-2008					
Number of observations (total)	2494	1508	986		
Number of industries (each year)	86	52	34		
Simple average across all observations					
TFP growth rate(%)	0.394	0.697	−0.069	3.626	0.000
	(5.173)	(5.473)	(4.641)		
Degree of competition (%)	99.064	94.134	106.606	−12.979	0.000
	(24.239)	(14.893)	(32.457)		
ΔDegree of competition (%)	0.044	0.151	−0.119	0.900	0.368
	(7.195)	(3.557)	(10.566)		
Ratio of R&D stock (%)	12.455	19.957	0.980	25.602	0.000
	(20.337)	(23.215)	(2.065)		
ΔRatio of R&D stock (%)	0.335	0.558	−0.008	6.616	0.000
	(2.070)	(2.633)	(0.209)		
Ratio of total IT investment (%)	2.378	1.355	3.943	−9.628	0.000
	(6.681)	(1.266)	(10.319)		
Ratio of household consumption (%)	23.870	18.761	31.684	−10.392	0.000
	(31.009)	(30.732)	(29.793)		
Ratio of export (%)	8.529	12.136	3.012	19.256	0.000
	(12.398)	(12.877)	(9.218)		

Table 4.4. Descriptive statistics.

	All	Manufacturing	Non-manufacturing	*t*-value	*p*-value
Year: 1980-1994					
Number of observations (total)	1290	780	510		
Number of industries (each year)	86	52	34		
Simple average across all observations					
TFP growth rate(%)	0.426	0.770	−0.100	2.722	0.007
	(5.627)	(5.844)	(5.238)		
Degree of competition (%)	99.913	93.674	109.455	−11.026	0.000
	(26.282)	(14.147)	(35.945)		
ΔDegree of competition (%)	−0.220	0.039	−0.615	1.428	0.153
	(7.767)	(3.430)	(11.598)		
Ratio of R&D stock (%)	9.589	15.212	0.988	19.095	0.000
	(14.812)	(16.722)	(2.243)		
ΔRatio of R&D stock (%)	0.352	0.593	−0.016	7.063	0.000
	(1.490)	(1.866)	(0.268)		
Ratio of total IT investment (%)	2.177	1.011	3.962	−7.523	0.000
	(7.034)	(1.155)	(10.862)		
Ratio of household consumption (%)	22.686	17.012	31.363	−9.014	0.000
	(28.816)	(26.757)	(29.705)		
Ratio of export (%)	7.462	10.593	2.672	13.735	0.000
	(10.840)	(10.740)	(9.111)		

Table 4.4. Descriptive statistics (continued).

	All	Manufacturing	Non-manufacturing	*t*-value	*p*-value
Year: 1995-2008					
Number of observations (total)	1204	728	476		
Number of industries (each year)	86	52	34		
Simple average across all observations					
TFP growth rate(%)	0.359	0.618	−0.037	2.400	0.017
	(4.640)	(5.048)	(3.907)		
Degree of competition (%)	98.155	94.626	103.553	−7.084	0.000
	(21.812)	(15.648)	(27.963)		
ΔDegree of competition (%)	0.405	0.349	0.491	−0.345	0.730
	(6.706)	(3.749)	(9.611)		
Ratio of R&D stock (%)	15.525	25.041	0.971	18.926	0.000
	(24.572)	(27.703)	(1.858)		
ΔRatio of R&D stock (%)	0.359	0.594	0.001	3.775	0.000
	(2.583)	(3.300)	(0.120)		
Ratio of total IT investment (%)	2.594	1.724	3.923	−6.028	0.000
	(6.277)	(1.276)	(9.714)		
Ratio of household consumption (%)	25.139	20.635	32.028	−5.910	0.000
	(33.164)	(34.407)	(29.916)		
Ratio of export (%)	9.672	13.789	3.376	13.782	0.000
	(13.788)	(14.656)	(9.328)		

Note: 1. The *t*-values represent statistics for the test of the difference in average between the manufacturing and non-manufacturing industries.
2. The standard deviations are reported in the round parentheses.
3. Some samples of the degree of competition (100 minus Lerner index) are over 100 because the Lerner index can be negative.

Table 4.4. Descriptive statistics (continued).

Notably enough, a positive change in the degree of competition in 1995-2008 appears to result from intensified competition both in the domestic and global markets. As Porter and Sakakibara (2004) aptly point out, deregulations and market liberalization in the domestic market since around mid-1990s (for example, the financial and retail industries) have grown competitors including foreign firms established by FDIs in Japan. As regards to the global market, the Japanese industries have been exposed to more fierce competition than ever before, encountering recent strong competitors, such as Korean and Chinese firms, through the modularization of production and the IT revolution in the same period. Thus, the change in the degree of competition after around mid-1990s should be interpreted as including both domestic and global competition effects.

4.5. Results

Tables 4.5 to 4.8 exhibit the results of the estimation formulated above. The dependent variable is the annual TFP growth rate. In addition to the independent variables listed in Table 4.1, year dummy variables are included in the formulation to control for business fluctuation in the economy, although they do not capture industrial business fluctuation. (Year dummy variables are omitted from the result tables to save space.) Two types of specifications are estimated: the first including only $comp$, and the second including both $comp$ and the quadratic term of $comp$ (i.e. $comp^2$) to affirm the inverted-U shape theory. Also, all industries are divided into the manufacturing and non-manufacturing industries in order to develop a detailed view of the underlying differences between the two industrial categories. Subsection 4.7.2 briefly discusses the results based upon the different classification in terms of the industries conducting or not conducting R&D investment. Moreover, Subsection 4.7.3 concisely conducts the first-difference (FD) estimation to confirm whether the competition effect is prevalent for the manufacturing and non-manufacturing industries.

The results of both the FE and FE with IVs estimations are presented, and robust (Eicker-Huber-White) clustered standard errors by industries are reported in all estimations. Although the FE (FE-IV) estimation in general posits that the industry fixed effects are constant over years, it seems somewhat hard to believe that the characteristics intrinsic to industries are unchanged over time, and in particular, this assumption may be difficult to hold during the entire sample period of 1980-2008. From this reason, the two subdivided periods of 1980-1994 and 1995-2008 are the main focus, where there seem to be fewer changes in the industry fixed effects. A note to be taken into account is the possibility that they may still change even in these subdivided periods, but it seems that industrial changes are relatively slow to occur as compared to those of firms due to the compound movements of firms within industries. In addition, the reason why we do not further subdivide the period is that this industry-level panel data has few observations in the cross-sectional dimension, and that if we confine the sample to a shorter period, we cannot obtain a sufficient number of observations in order for the estimation to attain any significant value. No matter what holds true, the general FE estimations using the whole sample require us to recognize the limitation, and hence, we treat them as a preliminary analysis.

4.5.1. Estimation results for the period of 1980-2008

Table 4.5 presents the results for the whole sample period: 1980-2008. Let us first look at the F-statistics of the exogenous test (Davidson-Mackinon test). This tests the null hypothesis that the FE-IV and FE estimations are both consistent, and the rejection implies the need for instrumenting. We can see that since almost all estimations except Estimation (7) reject the null hypothesis, the FE-IV estimations are largely more robust to inconsistency.

We also need to consider the underidentification test (Kleibergen-Paap rk LM test) and the weak identification test (Stock-Yogo test). The null hypothesis of the former test is that IVs are irrelevant, and the null hypothesis of the latter test is that IVs are weak against the alternative that they are strong. In particular, because the F-statistics of the weak identification test in Estimations (4), (8), and (12) (where we investigate the inverted-U shape theory) are fairly small, the finite-sample bias of these FE-IV estimations can be considerably large relative to those of the FE estimations. Consequently, we prefer the simple FE estimators for these estimations. Finally, the overidentification test (Hansen J test) in these estimations cannot reject the null hypothesis that all IVs are valid. Hereafter, the detailed interpretations of these tests are skipped for descriptive simplicity. [22]

With these in mind, we first examine whether increased competition indicates a rise in TFP growth or not in each industrial category. The coefficient of the degree

[22] Additionally, this study tests the null hypothesis that the unreported coefficients of year dummy variables are jointly zero. The test statistics, for example, of Estimation (3) are 4.56 ($p = 0.000$), and hence they are considered significant. Although the year dummy variables may not fully control business fluctuations or demand shocks, they are retained in all formulations.

of competition signifies a change in the TFP growth rate (% point) in response to a 1% point increase in the degree (level) of competition. Put differently from Equation (4.4), it estimates the effect of a 1% point decrease in industrial profitability on the TFP growth rate.

In view of all industries, Estimation (3) (FE-IV) shows a negative competition effect on the TFP growth rate, but the estimate is not significant at all ($p = 0.722$). Although the estimate of Estimation (1) (FE) is slightly positive, it is significant only at the 10% level ($p = 0.070$). Hence, it is unclear whether the degree of competition affects TFP growth in all industries during this sample period. In the manufacturing industries, both Estimations (5) (FE) and (7) (FE-IV) demonstrate that the one-year lag of the competition measure positively affects TFP growth at the 1% significance level. While Estimation (5) is preferred on the basis of the exogeneity test, the relevant coefficient still takes a positive value, 0.151. By contrast, it is revealed that in the non-manufacturing industries, the estimate of the competition measure is negative in Estimation (11) (FE-IV), but the significance level is over 10% ($p = 0.184$). On the other hand, although the estimate is positive in Estimation (9) (FE), it is not significant at the 10% level ($p = 0.167$), either. From these results, we have an approximate idea that while the degree of competition has a tendency to positively affect TFP growth in the manufacturing industries, its effect on TFP growth in the non-manufacturing industries are arguably ambiguous.

As for the inverted-U shape theory, Estimations (2) and (10) (the FE estimations of all and the non-manufacturing industries) suggest a very weak non-linear relation.

However, since we can hardly observe the inverted-U shapes in the other estimations, it does not seem that the inverted-U shape theory is considered so robust during this sample period.

We can also perceive that an increase in R&D stock ratio has positive effects on the non-manufacturing industries at the 1% significance level and on the manufacturing industries at the 5-10% significance level. Interestingly enough, the estimates of the non-manufacturing industries (about 3.64) are much larger than those of the manufacturing industries (about 0.22) judging from Estimations (7) and (11). Finally, it is demonstrated that the total IT investment ratio is not significant at all in every industrial category. [23] The later subsections (in particular, Subsection 4.5.4) attempt to provide explanations on how these results can be interpreted.

[23] In order to check the robustness of the analysis of the incremental R&D stock ratio and the total IT investment ratio, the estimations are conducted by excluding the “electricity" and “gas, heat supply" industries, which could be “outliers", from the observations of the non-manufacturing industries. But we find that the basic results are not subject to change.

Dependent variable: TFP growth rate

	All industries				Manufacturing industries				Non-manufacturing industries			
	(1) FE	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE	(6) FE	(7) FE-IV	(8) FE-IV	(9) FE	(10) FE	(11) FE-IV	(12) FE-IV
$comp_{it-1}$	0.040*	0.167***	−0.013	1.478	0.151***	−0.020	0.236***	−2.148	0.030	0.128**	−0.073	−0.105
	(0.022)	(0.041)	(0.036)	(1.227)	(0.031)	(0.164)	(0.085)	(2.098)	(0.021)	(0.048)	(0.055)	(0.709)
	[0.070]	[0.000]	[0.722]	[0.229]	[0.000]	[0.906]	[0.005]	[0.306]	[0.167]	[0.011]	[0.184]	[0.883]
$comp_{it-1}^2$		−0.0004***		−0.005		0.001		0.012		−0.0003**		0.0001
		(0.0001)		(0.004)		(0.001)		(0.011)		(0.0001)		(0.002)
		[0.000]		[0.263]		[0.302]		[0.247]		[0.014]		[0.964]
Δrds_{it-2}	0.278**	0.268**	0.292***	0.171	0.232*	0.233*	0.217**	0.246**	3.076***	3.011***	3.638***	3.666***
	(0.108)	(0.110)	(0.103)	(0.108)	(0.119)	(0.119)	(0.107)	(0.107)	(0.948)	(0.899)	(0.974)	(1.132)
	[0.012]	[0.017]	[0.005]	[0.113]	[0.057]	[0.056]	[0.043]	[0.022]	[0.003]	[0.002]	[0.000]	[0.001]
tit_{it-2}	−0.031	−0.032	0.054	0.026	−0.143	−0.165	−0.283	−0.484	−0.003	0.001	0.159	0.160
	(0.072)	(0.057)	(0.072)	(0.317)	(0.271)	(0.275)	(0.221)	(0.414)	(0.085)	(0.072)	(0.101)	(0.106)
	[0.666]	[0.584]	[0.455]	[0.935]	[0.600]	[0.551]	[0.201]	[0.243]	[0.969]	[0.985]	[0.117]	[0.132]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	2236	2236	2236	2236	1352	1352	1352	1352	884	884	884	884
# of indus.	86	86	86	86	52	52	52	52	34	34	34	34
F	4.902	5.724	4.802	5.344	9.235	9.158	9.533	9.120	42.228	81.347	31.378	27.722
Exogeneity test (Davidson-Mackinnon test)												
F			4.532	7.099			2.544	4.025			16.738	7.810
			[0.033]	[0.000]			[0.111]	[0.018]			[0.000]	[0.000]
Underidentification test (Kleibergen-Paap rk LM test)												
χ^2			17.519	3.175			24.061	2.553			8.621	0.521
			[0.001]	[0.366]			[0.000]	[0.466]			[0.035]	[0.914]
Weak identification test (Stock-Yogo test)												
F			88.306	3.013			142.625	4.281			29.616	0.563
Relative bias			<5%	>30%			<5%	>30%			<5%	>30%
Overidentification test (Hansen J test)												
χ^2			2.021	0.009			3.033	0.418			0.108	0.105
			[0.364]	[0.996]			[0.220]	[0.812]			[0.948]	[0.949]

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (3), (7), (11), and $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$, exp_{it-2} for (4), (8), (12).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.
4. The relative bias of the weak identification test represents the largest relative bias of the FE-IV estimators relative to the FE estimators.

Table 4.5. Estimation of the competition effect for the period of 1980-2008.

4.5.2. Estimation results for the period of 1980-1994

This now leads to the results obtained for the period of 1980-1994 presented in Table 4.6. In view of all industries, we cannot reject the null hypothesis of the exogeneity test at the conventional 5% level. When we have a look at the competition measure in Estimation (1) (FE), the estimate is positive at the 5% significance level ($p = 0.022$), but the numerical value, 0.051, is small. (This means that a 1% point increase in the degree of competition induces a rise in the TFP growth rate by 0.051%.) Next, in the manufacturing industries, the estimate of the competition measure is positive, 0.206, at the 1% significance level in Estimation (5) (FE), which is preferred rather than Estimation (7) (FE-IV) based upon the exogeneity test. We can also confirm that the relevant coefficient in Estimation (7), 0.219, is very close to the above value although it is significant at the 10% level ($p = 0.062$). On the other hand, Estimation (11) (FE-IV) in the non-manufacturing industries does not show any significant relation between the competition measure and the TFP growth rate ($p = 0.689$). Estimation (9) (FE), which is not preferred based upon the exogeneity test, provides a positive figure at the 5% significance level, but the estimate, 0.041, is small. Hence, we can see that while competition seems to improve TFP growth in the manufacturing industries, the competition effect is notably ambiguous in the non-manufacturing industries. (See the discussions in Subsection 4.5.4.)

Taking a look at the result of Estimation (2) (FE), we find that the competition measure and its quadratic term are significantly positive and negative, respectively, at the 1% significance level, which means that the inverted-U shape theory holds in all industries. Furthermore, it is demonstrated that in the non-manufacturing industries, the quadratic term of the competition measure is negative. But because the estimate of the quadratic term, -0.0004, is extremely small and significant at the 10% level ($p = 0.055$), the inverted-U shape does not seem strongly applicable in the non-manufacturing industries.

Whereas an increase in the R&D stock ratio has a positive effect on TFP growth in the non-manufacturing industries at the 1% significance level, we fail to find any significant effect in both all industries and the manufacturing industries. The total IT investment ratio is not significant in every industrial category, either.

Dependent variable: TFP growth rate

	All industries				Manufacturing industries				Non-manufacturing industries			
	(1) FE	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE	(6) FE	(7) FE-IV	(8) FE-IV	(9) FE	(10) FE	(11) FE-IV	(12) FE-IV
$comp_{it-1}$	0.051**	0.220***	−0.001	0.950	0.206***	−0.322	0.219*	−4.548	0.041**	0.174**	−0.015	0.107
	(0.022)	(0.065)	(0.030)	(0.775)	(0.044)	(0.395)	(0.117)	(4.213)	(0.018)	(0.079)	(0.039)	(0.322)
	[0.022]	[0.001]	[0.973]	[0.220]	[0.000]	[0.419]	[0.062]	[0.280]	[0.027]	[0.034]	[0.689]	[0.741]
$comp^2_{it-1}$		−0.001***		−0.003		0.003		0.024		−0.0004*		−0.0004
		(0.0002)		(0.002)		(0.002)		(0.021)		(0.0002)		(0.001)
		[0.002]		[0.246]		[0.171]		[0.254]		[0.055]		[0.687]
Δrds_{it-2}	0.266	0.269	0.278	0.290	0.258	0.267	0.258	0.338	4.377***	4.186***	4.616***	4.444***
	(0.241)	(0.249)	(0.231)	(0.288)	(0.303)	(0.305)	(0.299)	(0.294)	(1.418)	(1.146)	(1.615)	(1.364)
	[0.272]	[0.284]	[0.228]	[0.313]	[0.398]	[0.386]	[0.387]	[0.249]	[0.004]	[0.001]	[0.004]	[0.001]
tit_{it-2}	0.035	0.015	0.130	0.003	−0.084	−0.185	−0.104	−0.748	−0.016	−0.031	0.100	0.089
	(0.063)	(0.060)	(0.085)	(0.202)	(0.392)	(0.349)	(0.247)	(0.587)	(0.054)	(0.055)	(0.119)	(0.131)
	[0.578]	[0.805]	[0.127]	[0.990]	[0.832]	[0.597]	[0.675]	[0.202]	[0.775]	[0.572]	[0.402]	[0.501]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	1032	1032	1032	1032	624	624	624	624	408	408	408	408
# of indus.	86	86	86	86	52	52	52	52	34	34	34	34
F	3.902	5.645	3.282	3.264	10.368	11.883	4.004	5.150	7.678	11.248	4.777	6.033
Exogeneity test (Davidson-Mackinnon test)												
F			2.728	2.541			0.019	1.477			4.043	2.178
			[0.099]	[0.079]			[0.890]	[0.229]			[0.045]	[0.115]
Underidentification test (Kleibergen-Paap rk LM test)												
χ^2			12.941	3.146			14.473	2.197			8.415	4.099
			[0.005]	[0.370]			[0.002]	[0.533]			[0.038]	[0.251]
Weak identification test (Stock-Yogo test)												
F			84.097	2.789			71.727	2.554			33.085	2.431
Relative bias			<5%	>30%			<5%	>30%			<5%	>30%
Overidentification test (Hansen J test)												
χ^2			0.217	2.775			1.714	1.722			0.778	0.837
			[0.897]	[0.250]			[0.424]	[0.423]			[0.678]	[0.658]

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (3), (7), (11), and $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$, exp_{it-2} for (4), (8), (12).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.
4. The relative bias of the weak identification test represents the largest relative bias of the FE-IV estimators relative to the FE estimators.

Table 4.6. Estimation of the competition effect for the period of 1980-1994.

4.5.3. Estimation results for the period of 1995-2008

Finally, Table 4.7 indicates the results for the period of 1995-2008. As a composite effect, Estimation (1) (FE) in all industries demonstrates that the degree of competition has a slightly positive effect on TFP growth. In the manufacturing industries, we can confirm from Estimation (5) (FE) that not only is the competition effect still positive, 0.308, at the 1% significance level, but also the value gets larger than the estimated figure of 0.219 in 1980-1994. Moreover, Estimation (7) (FE-IV) also shows a positive value, 0.361, at the 1% significance level, so that the effect of the competition measure is rather robust. Hence, we can argue that the effect of competition in the manufacturing industries is likely to be observable throughout the all sample periods and that it gets stronger in the latter half of the sample than the former.

By contrast to the manufacturing industries, Estimation (11) (FE-IV) with regard to the non-manufacturing industries, which passes the exogeneity test, reveals that the degree of competition has a slightly negative effect on TFP growth (the coefficient is -0.087) at the 10% significance level ($p = 0.073$). Furthermore, the robustness-check estimation excluding the “electricity" and “gas, heat supply" industries from the sample shows that this coefficient of the competition effect decreases to -0.098 at approximately the 5% significance level ($p = 0.055$). From these findings, although the negative effect of competition is neither large nor highly significant, the Schumpeterian hypothesis may be seemingly applied in the non-manufacturing industries. (This result will be also carefully examined in Subsection 4.5.4.)

The inverted-U shape can be seen in all industries on the basis of Estimation (2)

(FE) in a similar fashion of the estimation in 1980-1995. However, there are no inverted-U shapes observed in the disaggregated estimations for the manufacturing and non-manufacturing industries.

As Estimation (5) (FE) indicates, the incremental R&D stock ratio in the manufacturing industries is significant at the 1% level by contrast to the estimation in 1980-1994. But the relevant estimates in the non-manufacturing industries are insignificant in 1995-2008. Consequently, the manufacturing and non-manufacturing industries may have faced during the mid-1990s some structural changes that operated oppositely for these two industrial categories. In addition, the total IT investment ratio is still insignificant in every industrial category.

Dependent variable: TFP growth rate

	All industries				Manufacturing industries				Non-manufacturing industries			
	(1) FE	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE	(6) FE	(7) FE-IV	(8) FE-IV	(9) FE	(10) FE	(11) FE-IV	(12) FE-IV
$comp_{it-1}$	0.079**	0.279**	0.090	0.397	0.308***	0.353	0.361***	−5.508	−0.003	−0.027	−0.087*	−0.794
	(0.036)	(0.109)	(0.056)	(0.954)	(0.061)	(0.378)	(0.103)	(6.702)	(0.018)	(0.091)	(0.049)	(0.644)
	[0.031]	[0.012]	[0.106]	[0.677]	[0.000]	[0.355]	[0.000]	[0.411]	[0.863]	[0.765]	[0.073]	[0.218]
$comp^2_{it-1}$		−0.001**		−0.001		−0.0002		0.030		0.0001		0.003
		(0.0004)		(0.004)		(0.002)		(0.034)		(0.0004)		(0.003)
		[0.029]		[0.770]		[0.902]		[0.380]		[0.784]		[0.296]
Δrds_{it-2}	0.325***	0.316***	0.323***	0.307***	0.282***	0.281***	0.275***	0.347***	2.493	2.509	3.426	3.670
	(0.098)	(0.098)	(0.094)	(0.110)	(0.101)	(0.101)	(0.092)	(0.127)	(2.922)	(2.919)	(2.445)	(2.511)
	[0.001]	[0.002]	[0.001]	[0.005]	[0.007]	[0.007]	[0.003]	[0.006]	[0.400]	[0.396]	[0.161]	[0.144]
tit_{it-2}	−0.062	0.005	−0.073	0.018	−0.489	−0.481	−0.570	−1.538	0.013	0.005	0.079	−0.174
	(0.043)	(0.059)	(0.063)	(0.429)	(0.331)	(0.348)	(0.366)	(1.050)	(0.041)	(0.051)	(0.055)	(0.268)
	[0.154]	[0.931]	[0.248]	[0.966]	[0.146]	[0.173]	[0.120]	[0.143]	[0.753]	[0.921]	[0.148]	[0.516]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	946	946	946	946	572	572	572	572	374	374	374	374
# of indus.	86	86	86	86	52	52	52	52	34	34	34	34
F	5.707	5.327	5.795	5.105	8.435	8.042	7.533	5.409	1.135	1.135	1.544	2.132
Exogeneity test (Davidson-Mackinnon test)												
F			0.088	0.236			0.849	3.734			3.979	3.702
			[0.767]	[0.790]			[0.357]	[0.025]			[0.047]	[0.026]
Underidentification test (Kleibergen-Paap rk LM test)												
χ^2			17.913	3.205			22.390	2.671			8.663	3.409
			[0.001]	[0.361]			[0.000]	[0.445]			[0.034]	[0.333]
Weak identification test (Stock-Yogo test)												
F			96.669	1.773			133.071	2.429			26.847	2.144
Relative bias			<5%	>30%			<5%	>30%			<5%	>30%
Overidentification test (Hansen J test)												
χ^2			4.116	9.680			0.747	0.321			1.547	0.349
			[0.128]	[0.008]			[0.688]	[0.852]			[0.461]	[0.840]

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (3), (7), (11), and $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$, exp_{it-2} for (4), (8), (12).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.
4. The relative bias of the weak identification test represents the largest relative bias of the FE-IV estimators relative to the FE estimators.

Table 4.7. Estimation of the competition effect for the period of 1995-2008.

In order to probe these control variables in detail, the model for the period between 2000 and 2008 is estimated, the result of which is shown in Table 4.8. It indicates that almost all the estimates of the incremental R&D stock ratio achieve a degree of significance and the numerical values become much larger, especially in the non-manufacturing industries. Accordingly, some structural changes are likely to have occurred in around the year 2000 again, interestingly, with R&D activity contributing to TFP growth. When it comes to Estimations (9) (FE) and (11) (FE-IV) in Table 4.8, the coefficients of the total IT investment ratio for the non-manufacturing industries are both positive, 0.312 ($p = 0.129$) and 0.471 ($p = 0.027$), respectively. [24] Thus, total IT investment may also have had a positive effect on TFP growth since around the year 2000. But this result does not seem much robust and needs to be further examined from various perspectives. (Indeed, the coefficients in Estimations [10] and [12] are not significant at all.) Subsection 4.5.4 below briefly discusses why the effects of R&D and IT investment changed in and around the year 2000.

---

[24] In the estimation that excludes the “electricity" and “gas, heat supply" industries from the sample, the relevant estimates of Estimations (9) and (11) are 0.351 ($p = 0.068$) and 0.533 ($p = 0.007$), respectively, which coefficients are larger and more significant than the estimation that uses the full sample of the non-manufacturing industries.

Dependent variable: TFP growth rate

	All industries				Manufacturing industries				Non-manufacturing industries			
	(1) FE	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE	(6) FE	(7) FE-IV	(8) FE-IV	(9) FE	(10) FE	(11) FE-IV	(12) FE-IV
$comp_{it-1}$	0.228***	0.276	0.272*	−4.158	0.399***	0.604	0.513***	−5.345	0.060	−0.080	−0.096	−2.994
	(0.069)	(0.243)	(0.150)	(8.078)	(0.108)	(0.737)	(0.134)	(5.335)	(0.048)	(0.207)	(0.078)	(8.559)
	[0.001]	[0.259]	[0.070]	[0.607]	[0.001]	[0.416]	[0.000]	[0.316]	[0.220]	[0.701]	[0.219]	[0.727]
$comp^2_{it-1}$		−0.0002		0.021		−0.001		0.030		0.001		0.013
		(0.001)		(0.038)		(0.004)		(0.027)		(0.001)		(0.038)
		[0.828]		[0.585]		[0.776]		[0.273]		[0.496]		[0.728]
Δrds_{it-2}	0.362***	0.361***	0.352***	0.476*	0.313***	0.310***	0.291***	0.400***	11.182***	11.299***	11.590***	14.173
	(0.072)	(0.071)	(0.068)	(0.265)	(0.079)	(0.079)	(0.072)	(0.122)	(3.363)	(3.402)	(3.408)	(9.020)
	[0.000]	[0.000]	[0.000]	[0.073]	[0.000]	[0.000]	[0.000]	[0.001]	[0.002]	[0.002]	[0.001]	[0.116]
tit_{it-2}	0.015	0.034	−0.030	−1.579	−0.956	−0.913	−1.106	−2.331	0.312	0.258	0.471**	−0.525
	(0.292)	(0.353)	(0.335)	(2.899)	(0.848)	(0.891)	(0.869)	(1.494)	(0.200)	(0.260)	(0.213)	(2.654)
	[0.959]	[0.924]	[0.929]	[0.586]	[0.265]	[0.310]	[0.203]	[0.119]	[0.129]	[0.330]	[0.027]	[0.843]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	516	516	516	516	312	312	312	312	204	204	204	204
# of indus.	86	86	86	86	52	52	52	52	34	34	34	34
F	11.652	10.738	8.871	7.247	12.004	10.800	7.706	5.054	3.154	6.944	2.982	2.762
Exogeneity test (Davidson-Mackinnon test)												
F			0.236	0.457			1.686	4.137			2.103	1.018
			[0.627]	[0.633]			[0.195]	[0.017]			[0.149]	[0.364]
Underidentification test (Kleibergen-Paap rk LM test)												
χ^2			15.778	0.589			18.999	3.827			4.130	0.183
			[0.001]	[0.899]			[0.000]	[0.281]			[0.248]	[0.980]
Weak identification test (Stock-Yogo test)												
F			45.555	0.200			79.161	3.776			10.889	0.067
Relative bias			<5%	>30%			<5%	>30%			5-10%	>30%
Overidentification test (Hansen J test)												
χ^2			0.363	1.171			0.924	0.514			1.744	0.587
			[0.834]	[0.557]			[0.630]	[0.773]			[0.418]	[0.746]

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (3), (7), (11), and $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$, exp_{it-2} for (4), (8), (12).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.
4. The relative bias of the weak identification test represents the largest relative bias of the FE-IV estimators relative to the FE estimators.

Table 4.8. Estimation of the competition effect for the period of 2000-2008.

4.5.4. Summary and discussions

The results derived from the above-mentioned analyses are summarized in what follows.

1. The competition effect on TFP growth in the manufacturing industries is positive over all the sample periods, 1980-2008, and it gets larger and more robust in the latter half of the sample period, 1995-2008. On the other hand, the competition effect in the non-manufacturing industries may be slightly negative in 1995-2008, which suggests that the Schumpeterian hypothesis can be applied in this industrial category. As a result, the composite competition effect in all industries is slightly positive both in 1980-1994 and 1995-2008.

2. There is a weak inverted-U shape relationship between the competition measure and TFP growth observed in almost all industries at the aggregated level.

3. While the incremental R&D stock ratio is significant for the non-manufacturing industries in 1980-1994, it is also the same for the manufacturing industries in 1995-2008. The estimates of the non-manufacturing industries are much larger than those of the manufacturing industries. All the estimates including both the manufacturing and non-manufacturing industries in 2000-2008 are significant and become larger than in the previous periods.

4. The total IT investment ratio of the non-manufacturing industries may have had a positive effect on the TFP growth rate in 2000-2008.

Result 1, that competition positively affects TFP growth in the manufacturing industries is consistent with many previous studies (Nickell, 1996; Okada, 2005; Inui et al., 2012). We could imagine that certain sectors within the manufacturing industries, which gradually expanded to the global market, may be subject to strong competitive pressures at both domestic and global levels, and thereby, this result may be drawn. On the other hand, it contradicts with that derived by Flath (2011) who insists that there is no relation between the Lerner index (price-cost margin) and innovation at the industry level. The difference in the results between these two studies seems to lie in what follows. While Flath (2011) constructs the cross-sectional industry-level data and uses a time-average competition measure, this study adopts the industry-specific panel data and employs the one-year lag of the competition measure as an independent variable based on the assumption that the effect of competition in the industry level has a time lag. It is not always easily determined as to which study is definitely credible, but this study stands to be more plausible as it takes into consideration both the possibility of reverse causality and industry-specific fixed effects in panel data, which generally have richer information than cross-section data. (Flath [2011]'s number of observations totals only 74.)

We also discover a new result that competition may have negatively affected TFP growth in the non-manufacturing industries between 1995 and 2008, which no researchers have comprehensively examined so far with a specific focus on this industrial category. How can these results be interpreted and what policy implications can be derived? At the same time, what we have to take into account is that the inverted-U shape cannot be robustly observed in the non-manufacturing industries. While Aghion et al. (2005) point to the Schumpeterian effect such that increased competition would lower productivity growth only when the degree of competition is already sufficiently high, our finding does not indicate such a non-linear relation but instead shows a simple linear negative (though moderate) relation between competition and TFP growth.

From a practical viewpoint, our estimation results would seemingly cause a serious concern in terms of competition policies since there seems to be hardly any leeway to raise industrial productivity by promoting further competition. Specifically, one may be tempted to inadvertently conclude that overcompetition would be the main cause of stagnating TFP growth in the non-manufacturing industries and that eliminating overcompetition could provide benefits to the relevant industries. However, we need to be carefully enough to derive direct and strong policy implications from our finding due to the model specification and remaining endogeneity problems. Hence, it is more prudent to first relate our finding to existing studies that explore factors leading to the relatively low productivity of the Japanese non-manufacturing industries, which in a way would lend support to our research. What follows intends to briefly introduce the possible reasons that seem associated with the negative competition effect on TFP growth.

In the first place, the empirical observation that there is a contrasting difference in the competition effect between the manufacturing and non-manufacturing industries especially in 1995-2008 may be interpreted by the Schumpeterian growth theory proposed by Aghion et al. (2009). Their theory presents that whereas the threat of entry by frontier firms induces incumbent firms in sectors that are close to the technology frontier to further innovate in hope of offsetting entry by frontier firms, it deteriorates incumbent firms' incentives to innovate in sectors that lag far behind the technology frontier. In our analysis, the simple averaged TFP levels in the manufacturing and non-manufacturing industries, being normalized to 100 in 1980, indicate 131.0 and 102.6, respectively, in 2008. It is thus suggested that incumbent firms in the non-manufacturing industries which seem to be laggard behind the technology frontier may have relatively weaker incentives to spur innovation, as compared to the manufacturing industries.

Second, regulations in the non-manufacturing industries could hinder competition in the industries from increasing TFP growth. According to the Cabinet Office of Japan (2006), indeed, regulatory reforms have progressed steadily in the manufacturing industries rather than the non-manufacturing industries. [25] As many economists have pointed out, further regulatory reforms in the non-manufacturing industries need to be effectively implemented in such a manner to make the competition effect work accordingly, likewise in the manufacturing industries, and thus to raise TFP growth (Nakanish and Inui, 2007). [26]

[25] The Cabinet Office of Japan (2006) reports the regulation index that indicates the progress of regulatory reforms in each industry based on the classification of the *JIP Database* covering 1995-2005. The regulation index is normalized to 1 in 1995 if regulation exists in a particular industrial sector, and the closer the numerical value is to 0, the further regulatory reforms advance in comparison to 1995. The simple average values of this regulation index in the non-manufacturing industries in 1995 and 2008 are 0.971 and 0.600, while they are 0.635 and 0.273 in the manufacturing industries. Hence, more regulations have continued to exist in the non-manufacturing industries.

[26] Nakanishi and Inui (2007) investigate the effects of regulations using the *JIP Database* between 1970 and 2002. The measure is based on the regulation index indicated in Footnote 25. They uncover that regulations in industries not conducting R&D investment, which are mostly the non-manufacturing industries, have a negative effect on both TFP and production growth. Similarly, this study attempts to conduct a regression analysis in the available period of 1995-2005 by adding the one-year lag of the regulation measure as an independent variable instead of the one-year lag of the competition measure.

The third reason is that although they are not entailed in the rigorous definition of the Schumpeterian hypothesis and innovation, the distribution, entry, and exit effects mentioned in Section 4.4 are too weak to support the positive competition effect in the non-manufacturing industries. In particular, Kim et al. (2010) indicate that the exit effect in the non-manufacturing industries was consistently negative between 1980 and 2005. Then, the authors argue that highly productive firms fail to increase their market shares and that due to this absence of refreshment in industries, the stagnation of the non-manufacturing industries dampens TFP growth. In addition, Morikawa (2011), who uses the Japanese firm-level micro data, finds the empirical observation that in view of service industries (nearly non-manufacturing industries), the contributions of the distribution, entry, and exit to the productivity are negative, and especially those of retail industries are small. Hence, it can be expected that if the exit of low-productivity firms, mainly in service industries, promoted by competition works properly, the industrial TFP growth rate will rise.

From these studies, it seems more important in the non-manufacturing industries to encourage sound competition to work through the measures such as regulatory reforms and industrial refreshment policies than to stifle competition simply based on the observation about the linear negative relationship. But we should realize that it is possible only to speculate these reasons for the application of the Schumpeterian hypothesis

(The IV method is not utilized.) According to this analysis, the estimate of the regulation index is significantly negative ($p = 0.053$) in the manufacturing industries, but it is not significant in the least in the non-manufacturing industries, which derives a different result from that of Nakanishi and Inui (2007). In our study, it is therefore concluded that whereas removal of regulations positively affected TFP growth in the manufacturing industries possibly through intensified market competition, deregulation in the non-manufacturing industries ceased to work functionally as well as the progress was unsatisfactory.

and the conceivable prescriptions for vitalizing the effect of competition in the non-manufacturing industries. In particular, since each sector in the non-manufacturing industries (e.g. wholesale, retail, transportation, accommodation, among others) is quite heterogeneous in many aspects, it might be incorrect to deal with these industries as if they are a “one sector" and to propose uniform policy recommendations neglecting detailed characteristics in each industrial sector. [27] Furthermore, Morikawa (2014) reveals that the distribution of TFP among firm levels in the service industries is more widely varied than in the manufacturing industries. Hence, further examination is required to uphold the validity of these claims being made, as briefly mentioned in Section 4.6.

Result 2 is somewhat different from other studies such as Inui et al. (2012) who find the inverted-U shape relation between the degree of competition and productivity using the data of manufacturing industries. One possible reason why this study displays a weak inverted-U shape in all industries, but not in other industrial categories, would be that: whereas the firm-specific data used in existing studies has ample observations being broadly distributed along with the competition measure and productivity growth as the theory predicts, the industry-specific data is accumulated from such a small number observations that we cannot detect a clear-cut relation, because observations undergo little changes due to the rigidity of movements at the industrial level.

[27] A problem also arises in that TFP cannot perfectly capture real productivity in service industries (nearly manufacturing industries). More precisely, although customer satisfaction is the most important feature if service sectors in Japan, TFP is difficult to measure such quality.

With regard to Result 3, we are left pondering about the difference in the impacts of an increase in R&D stock ratio between the manufacturing (significant in 1995-2008) and non-manufacturing industries (significant in 1980-1994). [28] One possible interpretation for the manufacturing industries could be that because these industries were not subjected to intense competition in the former sample period in comparison with concurrent global competitive environments, they may not have sensed a need to conduct R&D which aims at immediately obtaining productivity improvement. But since the latter half of the 1990s, the manufacturing industries engaged in global competition may have seemingly felt required to conduct conducive R&D, which could be more focused on development and application. [29] On the other hand, the R&D stock to output ratio of the non-manufacturing industries is significantly smaller than that of the manufacturing industries (Table 4.4), so that the impact of an increase in the ratio on TFP growth tends to be strikingly large. While an average decrease in the R&D stock ratio in 1980-1994 (−0.016% point) contributed to negative TFP growth (namely, the coefficient is positive), the minuscule increase of that ratio (0.001% point) in 1995-2008 would not afford to have a significant positive impact on TFP growth.

Finally, the finding of Result 3 showing that the effect of an increase in R&D is larger and significant in every industrial category in 2000-2008 may reflect an IT advancement exemplified by the Internet, which has been widespread throughout in-

[28] Kwon et al. (2010), who employ both the business finance data released by the Development Bank of Japan and the R&D data of the *Report on the Survey of Research and Development* produced by NISTEP, find that R&D intensity positively affects TFP growth irrespective of time periods and industries, which result differs from that derived in this chapter.

[29] Kwon et al. (2010) also prove the larger and more stable effect of development and application research on TFP growth than basic research conducted by Japanese firms.

dustries since the beginning of this century. As Result 4 suggests, there is also a possibility that IT may enhance the effectiveness of R&D conducted by the Japanese industries. [30] On the other hand, the effect of total IT investment is significantly positive solely on the non-manufacturing industries. Although it appears paradoxical, a possible reason is that while the non-manufacturing industries such as finance have made better use of IT, the manufacturing industries may not have fully realized the beneficial utilization of it. [31]

The above-mentioned arguments of Results 3 and 4 are mere conjectures, not necessarily based on concrete data and evidence. Accordingly, the underlying economic mechanism should be further scrutinized by using a micro firm-level database rather than the industry-level *JIP Database*.

4.6. Concluding remarks

This chapter mainly investigated the causal relationship between the degree of competition, which is measured by the Lerner index, and the TFP growth rate on the basis of the Japanese industry-level panel data (the *JIP Database*) from 1980 to 2008. The central finding indicated that although a positive effect of competition on the TFP growth rate is clearly observable in the manufacturing industries throughout the sample pe-

[30] In spite of the old Japanese industry-level dataset from 1980-1998, Nishimura and Shirai (2003) discover the existence of the positive externality effect of information and communication technology on the technological progress (they term it as “New Economy effect") in the manufacturing industries.

[31] The Ministry of Economy, Trade and Industry of Japan (2015) reports not only that Japanese manufacturing industries fall behind foreign companies in terms of the application of IT, but also that their IT investments are mostly concentrated on operational efficiency improvement and cost reduction, not on innovative activities that can be expected to result in greater TFP growth.

riod (1980-2008), such effect in the non-manufacturing industries may be slightly negative in the latter half of the sample period (1995-2008). This finding of a negative competition effect may lend support to the claim that the Schumpeterian hypothesis can be applied in the case of the non-manufacturing industries. Furthermore, a weak inverted-U shape relation between the competition measure and TFP growth proposed by Aghion et al. (2005) can be limitedly seen in all industries. An increase in the R&D stock ratio stimulates TFP growth of the manufacturing industries in 1995-2008 and of the non-manufacturing industries in 1980-1995. However, we cannot observe any significant relation between the total IT investment ratio and TFP growth except for the non-manufacturing industries in 2000-2008.

As already mentioned, we must bear in mind that even if the Schumpeterian hypothesis seems applicable to the non-manufacturing industries, it never derives the simple conclusion that the market structures limiting competition, such as monopoly, are unequivocally desirable for productivity improvements and innovative activities in these industries. As the standard microeconomics theory indicates, monopoly usually causes inefficient resource allocation in the form of deadweight loss and transfers a portion of consumer surplus to producers. Furthermore, monopolists sometimes devote an abundance of energy to rent-seeking activity in order to maintain their current monopoly rents and to exclude potential rivals from the markets. Indeed, certain monopoly power may have to be approved, for example, by awarding patents to firms that have innovated to compensate them for their R&D costs and efforts, but it is also of more importance to eliminate obstacles that prevent market competition from interacting well with productivity improvements, such as unnecessary regulations and stagnation of exit and entry in the industries as referred in Subsection 4.5.4. It can be therefore potentially misleading to conclusively decide the course of action that competition or innovation policies should take toward restricting sound competition in the non-manufacturing industries based solely upon the results of this study. Nevertheless, it is reasonable to believe that allowing competition to work harmoniously in the non-manufacturing industries is the key to raising their productivity, and thereby, restoring the Japanese economy to growth.

The present study is subject to further debate. First, it needs to be proven whether the result that competition may imply a negative effect on TFP growth in the non-manufacturing industries (the Schumpeterian hypothesis) is valid or not by using firm-level micro data. In this regard, it is strongly desired to build firm-level datasets that allows for such an analysis to be carried out. In addition, another direction of the study is to identify how competition policies or regulatory reforms affect competitive environments and productivity by highlighting a natural experiment such as a policy change (e.g. difference-in-difference approach) on the basis of firm-level data, which would provide a basis for a policy prescription that aims at improving productivity. Second, although this study regards some control variables as exogenous, the model specification can be further improved by implementing, for example, simultaneous equation models. Moreover, a more suggestive estimation framework may be needed if there exists an intention to examine the impact of factors such as trade openness and production networks by using data of import penetration, intra-industry trade, and FDIs. Finally, assuming that competition has some effects on productivity of industries, whether they are positive or negative, we need to conduct further study on the detailed mechanism in force within them, such as how and the extent to which innovative activities react to incentives.

4.7. Appendices

4.7.1. Industry-specific Lerner index

Existing studies such as Nickell (1996), Okada (2005), and Aghion et al. (2005) calculate the average Lerner index of firms, LI_F, as follows:

$$LI_F = \frac{sales - cost\ of\ sales + depreciation - rK}{sales}, \tag{4.5}$$

where r is the cost of capital and K is the capital stock. According to Fukao et al. (2008), the industry-specific Lerner index, LI_I, can be formally defined as follows:

$$LI_I = \frac{\Psi}{p_Q Q} = \frac{p_Q Q - p_M X - wL - rK}{p_Q Q}, \tag{4.6}$$

where Ψ represents monopoly rents, and X, L, and K are the total amounts of intermediate, labor, capital service inputs, respectively. In addition, p_Q, p_M, w, and r denote the market prices for final output and intermediate input, wage rate, and capital costs, respectively.

Because gross output measured by the factor costs is equivalent to the sum of the intermediate input, compensation of employment, operating profits, and consumption of fixed capital, the nominator in Equation (4.6) (monopoly rent) should equal the sum of the operating profits and compensation of fixed capital minus the capital service input. Taking into account that the above operating profits corresponds to “sales − cost of sales" in Equation (4.5), we see that the interpretation of LI_I is the same as the average Lerner index of firms, LI_F. Hence, firms belonging to that industry are expected to gain average profits in proportion to LI_I.

It should be noted that the above formulation of Equation (4.6) has been derived on the basis of the several simple assumptions. For example, it is assumed that perfect competition prevails in the factor production markets and the markup is constant over time. It would be therefore more feasible to regard the industry-specific Lerner index as a proxy for market power rather than accurate profitability.

4.7.2. Estimation for industries conducting and not conducting R&D investment

This subsection performs the same regression analysis as before by dividing all industries into the industries conducting and not conducting R&D investment. While the 75 industries, manufacturing industries constituting a majority and the remainder non-manufacturing industries, conduct R&D investment, the 11 industries, for example, eating/drinking places and accommodation, do not invest in R&D. A note worth mentioning is that because the number of observations of the industries not conducting R&D investment are fairly limited, we need to interpret the following results in a careful manner.

Table 4.9 presents the results obtained for 1980-1994. For industries conducting R&D investment, Estimation (7) (FE-IV), which passes the exogeneity test, demonstrates that the effect of the competition measure is negative but not significant at all. On the other hand, Estimation (5) (FE) shows a significantly positive effect on TFP growth, but the numerical value, 0.092, is small and close to zero. With regard to the industries not conducting R&D investment, the competition effects are hardly observed, although Estimation (11) (FE-IV) exhibits a slightly positive coefficient, 0.042, at the 10% significance level ($p = 0.057$). Furthermore, surprisingly enough, we can find that the increased R&D stock ratio has no significant effects on TFP growth in this sample period.

Let us turn to Table 4.10 that presents the results for 1995-2008. Contrasting to the previous results, it reveals from Estimations (5) (FE) and (7) (FE-IV) that increased competition has a significant larger impact on TFP growth in the industries conducting R&D investment. But for the industries not conducting R&D investment, the competition measures in Estimations (9) (FE) and (11) (FE-IV) are not significant while the signs of the coefficients are slightly negative. It is also shown that an increase in the R&D stock ratio positively affects TFP growth in this latter half of the sample period. Lastly, the total IT investment ratio may have a positive impact on TFP growth for the industries not conducting R&D, the estimates in Estimations (9) and (11) being significant at around the 15% ($p = 0.159$) and 1% ($p = 0.011$) significance levels, respectively. This could suggest that the application of IT may be better promoted especially in industries not conducting R&D. However, we need to note that the estimates of the total IT investment ratio are not robust, as Estimation (5) in the industries conducting R&D investment exhibits a highly significant and negative effect. Hence, the results concerning this variable should be carefully evaluated.

In conclusion, it seems that the industries conducting R&D investment not only improve their industrial productivity by being exposed to competition, but also make R&D investment more effective for the period of 1995-2008, namely after the bubble economy burst.

Dependent variable: TFP growth rate

	All industries				Industries conducting R&D				Industries not conducting R&D			
	(1) FE	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE	(6) FE	(7) FE-IV	(8) FE-IV	(9) FE	(10) FE	(11) FE-IV	(12) FE-IV
$comp_{it-1}$	0.051**	0.220***	−0.001	0.950	0.092**	0.294***	−0.023	3.987	0.002	0.119	0.042*	0.289
	(0.022)	(0.065)	(0.030)	(0.775)	(0.040)	(0.111)	(0.069)	(3.416)	(0.009)	(0.087)	(0.022)	(0.358)
	[0.022]	[0.001]	[0.973]	[0.220]	[0.023]	[0.010]	[0.791]	[0.389]	[0.821]	[0.203]	[0.057]	[0.420]
$comp^2_{it-1}$		−0.001***		−0.003		−0.001**		−0.015		−0.0003		−0.001
		(0.0002)		(0.002)		(0.0003)		(0.014)		(0.0002)		(0.001)
		[0.002]		[0.246]		[0.027]		[0.271]		[0.166]		[0.516]
Δrds_{it-2}	0.266	0.269	0.278	0.290	0.284	0.284	0.296	0.307				
	(0.241)	(0.249)	(0.231)	(0.288)	(0.250)	(0.257)	(0.230)	(0.401)				
	[0.272]	[0.284]	[0.228]	[0.313]	[0.260]	[0.272]	[0.198]	[0.444]				
tit_{it-2}	0.035	0.015	0.130	0.003	−0.022	0.026	0.176	1.150	0.183	0.031	−0.346	−0.737
	(0.063)	(0.060)	(0.085)	(0.202)	(0.094)	(0.070)	(0.129)	(1.633)	(0.659)	(0.735)	(0.740)	(1.024)
	[0.578]	[0.805]	[0.127]	[0.990]	[0.813]	[0.715]	[0.172]	[0.481]	[0.787]	[0.968]	[0.640]	[0.472]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	1032	1032	1032	1032	900	900	900	900	143	143	132	132
# of indus.	86	86	86	86	75	75	75	75	11	11	11	11
F	3.902	5.645	3.282	3.264	4.143	5.007	2.759	2.729	–	–	16.136	31.601
Exogeneity test (Davidson-Mackinnon test)												
F			2.728	2.541			4.016	5.818			0.756	1.021
			[0.099]	[0.079]			[0.045]	[0.003]			[0.386]	[0.364]
Underidentification test (Kleibergen-Paap rk LM test)												
χ^2			12.941	3.146			12.234	1.039			2.480	1.477
			[0.005]	[0.370]			[0.007]	[0.792]			[0.479]	[0.688]
Weak identification test (Stock-Yogo test)												
F			84.097	2.789			52.294	0.870			16.734	1.210
Relative bias			5<%	>30%			5<%	>30%			5<%	>30%
Overidentification test (Hansen J test)												
χ^2			0.217	2.775			1.165	0.606			–	–
			[0.897]	[0.250]			[0.559]	[0.739]			–	–

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (3), (7), (11), and $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$, exp_{it-2} for (4), (8), (12).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.
4. The relative bias of the weak identification test represents the largest relative bias of the FE-IV estimators relative to the FE estimators.

Table 4.9. Estimation of the competition effect for the period of 1980-1994 (industries conducting and not conducting R&D).

Dependent variable: TFP growth rate

	All industries				Industries conducting R&D				Industries not conducting R&D			
	(1) FE	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE	(6) FE	(7) FE-IV	(8) FE-IV	(9) FE	(10) FE	(11) FE-IV	(12) FE-IV
$comp_{it-1}$	0.079**	0.279**	0.090	0.397	0.173***	0.418***	0.156**	1.541	−0.017	−0.274**	−0.095	−0.872
	(0.036)	(0.109)	(0.056)	(0.954)	(0.048)	(0.131)	(0.074)	(3.106)	(0.030)	(0.097)	(0.068)	(0.773)
	[0.031]	[0.012]	[0.106]	[0.677]	[0.001]	[0.002]	[0.035]	[0.620]	[0.569]	[0.018]	[0.162]	[0.259]
$comp^2_{it-1}$		−0.001**		−0.001		−0.001**		−0.007		0.001**		0.003
		(0.0004)		(0.004)		(0.001)		(0.015)		(0.0003)		(0.003)
		[0.029]		[0.770]		[0.034]		[0.665]		[0.019]		[0.276]
Δrds_{it-2}	0.325***	0.316***	0.323***	0.307***	0.302***	0.294***	0.305***	0.254*				
	(0.098)	(0.098)	(0.094)	(0.110)	(0.098)	(0.098)	(0.092)	(0.155)				
	[0.001]	[0.002]	[0.001]	[0.005]	[0.003]	[0.004]	[0.001]	[0.100]				
tit_{it-2}	−0.062	0.005	−0.073	0.018	−0.159***	−0.030	−0.144***	0.531	1.324	1.498*	1.775***	2.107***
	(0.043)	(0.059)	(0.063)	(0.429)	(0.056)	(0.101)	(0.073)	(1.725)	(0.871)	(0.759)	(0.696)	(0.668)
	[0.154]	[0.931]	[0.248]	[0.966]	[0.005]	[0.771]	[0.049]	[0.758]	[0.159]	[0.077]	[0.011]	[0.002]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	946	946	946	946	825	825	825	825	132	132	121	121
# of indus.	86	86	86	86	75	75	75	75	11	11	11	11
F	5.707	5.327	5.795	5.105	6.524	6.442	5.908	5.724	–	–	1.172	0.810
Exogeneity test (Davidson-Mackinnon test)												
F			0.088	0.236			0.164	0.116			1.556	1.029
			[0.767]	[0.790]			[0.686]	[0.891]			[0.215]	[0.361]
Underidentification test (Kleibergen-Paap rk LM test)												
χ^2			17.913	3.205			33.099	1.135			4.641	4.847
			[0.001]	[0.361]			[0.000]	[0.769]			[0.200]	[0.183]
Weak identification test (Stock-Yogo test)												
F			96.669	1.773			168.587	0.406			5.786	0.594
Relative bias			<5%	>30%			<5%	>30%			20–30%	>30%
Overidentification test (Hansen J test)												
χ^2			4.116	9.680			0.918	5.406			–	–
			[0.128]	[0.008]			[0.632]	[0.067]			–	–

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (3), (7), (11), and $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$, exp_{it-2} for (4), (8), (12).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.
4. The relative bias of the weak identification test represents the largest relative bias of the FE-IV estimators relative to the FE estimators.

Table 4.10. Estimation of the competition effect for the period of 1995-2008 (industries conducting and not conducting R&D).

4.7.3. First-difference estimation

In addition to the FE model, the FD model is a possible alternative method toward eliminating unobservable fixed effects. By differencing both sides of Equation (4.2) between t and $t-1$, we obtain the FD estimation model as follows:

$$\Delta tfpg_{it} = \Delta\alpha_t + \beta_1 \Delta comp_{it-1} + \beta_2 \Delta(\Delta rds_{it-2}) + \beta_3 \Delta tit_{it-2} + \Delta\varepsilon_{it}. \qquad (4.7)$$

The FE and FD estimators are both unbiased for the underlying formulation. Moreover, although both estimators are consistent with fixed T as $I \to \infty$, our dataset with small I is expected to produce different estimates between FE and FD estimation.

While it is sometimes difficult to opt for one estimation method over the other, the FD estimator is more efficient than the FE estimator if the first difference of the error term, $e_{it} \equiv \Delta\varepsilon_{it}$ for $t = 2, \cdots T$, is serially uncorrelated. (That is, no serial correlation of e_{it} means ε_{it} being subject to a random walk.) We can simply test for a serial correlation of e_{it} using the method proposed by Wooldridge (2010), [32] but unfortunately, the test proves that e_{it} is significantly serially correlated for all types of estimations. Furthermore, it is demonstrated that the inconsistency in FE estimators is smaller than FD estimators if T is reasonably large under contemporaneous exogeneity. (See Footnote 10.) For these reasons, there seems to be no positive rationale for employing the FD estimator, so that this subsection presents the following estimation results only as a reference.

[32] According to Wooldridge (2010), a test can be conducted for serial correlation of the error term in the first-differenced Equation (4.7) by regressing a pooled OLS model as for the formulation, $\hat{e}_{it} = \hat{\rho}\hat{e}_{it-1} + error_{it}$ ($t = 3, 4, \ldots, T$ and $i = 1, 2, \ldots, I$), and performing t-test on a $\hat{\rho}$.

Dependent variable: TFP growth rate

	All		Manufacturing		Non-manufacturing	
	(1)	(2)	(3)	(4)	(5)	(6)
	FD	FD-IV	FD	FD-IV	FD	FD-IV
1980-1994	0.195*	0.177*	1.016***	0.919***	0.090*	0.067
	(0.105)	(0.103)	(0.233)	(0.207)	(0.049)	(0.045)
	[0.066]	[0.086]	[0.000]	[0.000]	[0.077]	[0.136]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	946	946	572	572	374	374
# of indus.	86	86	52	52	34	34
F	4.397	4.244	7.891	7.607	22.088	20.943
Exogeneity test (Davidson-Mackinnon test)						
F		4.697		3.795		8.420
		[0.033]		[0.057]		[0.007]
Underidentification test (Kleibergen Paap rk LM test)						
χ^2		12.271		19.102		9.485
		[0.007]		[0.000]		[0.024]
Weak identification test (Stock-Yogo test)						
F		587.991		552.202		218.963
Relative bias		<5%		<5%		<5%
Overidentification test (Hansen *J* test)						
χ^2		5.573		4.258		0.687
		[0.062]		[0.119]		[0.709]
1995-2008	0.499***	0.489***	1.193***	1.227***	0.152**	0.106
	(0.132)	(0.142)	(0.084)	(0.105)	(0.069)	(0.066)
	[0.000]	[0.001]	[0.000]	[0.000]	[0.035]	[0.109]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	860	860	520	520	340	340
# of indus.	86	86	52	52	34	34
F	7.190	6.753	34.911	30.767	4.862	3.912
Exogeneity test (Davidson-Mackinnon test)						
F		0.123		0.376		3.899
		[0.727]		[0.543]		[0.057]
Underidentification test (Kleibergen Paap rk LM test)						
χ^2		15.230		20.972		7.578
		[0.002]		[0.000]		[0.056]
Weak identification test (Stock-Yogo test)						
F		743.523		566.691		253.055
Relative bias		<5%		<5%		<5%
Overidentification test (Hansen *J* test)						
χ^2		1.659		1.397		4.668
		[0.436]		[0.497]		[0.097]

Table 4.11. First-difference estimation of the competition effect.

Dependent variable: TFP growth rate

	All		Manufacturing		Non-manufacturing	
	(1)	(2)	(3)	(4)	(5)	(6)
	FD	FD-IV	FD	FD-IV	FD	FD-IV
2000-2008	0.527***	0.623**	1.130***	1.277***	0.117	0.078
	(0.186)	(0.265)	(0.149)	(0.191)	(0.112)	(0.138)
	[0.006]	[0.019]	[0.000]	[0.000]	[0.303]	[0.572]
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
# of obs.	430	430	260	260	170	170
# of indus.	86	86	52	52	34	34
F	5.026	4.370	13.925	11.194	2.002	1.785
Exogeneity test (Davidson-Mackinnon test)						
F		1.810		2.424		0.408
		[0.182]		[0.126]		[0.528]
Underidentification test (Kleibergen Paap rk LM test)						
χ^2		25.383		23.188		9.050
		[0.000]		[0.000]		[0.029]
Weak identification test (Stock-Yogo test)						
F		171.026		196.282		48.154
Relative bias		<5%		<5%		<5%
Overidentification test (Hansen J test)						
χ^2		1.114		0.159		0.301
		[0.573]		[0.923]		[0.860]

Note: 1. $\Delta comp_{it-1}$, $\Delta comp_{it-2}$, $hcons_{it-2}$ are used as IVs for (2), (4), (6).
2. ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.
3. The robust clustered standard errors and p-values are reported in the round and square parentheses, respectively.

Table 4.11. First-difference estimation of the competition effect (continued).

The FD estimates of the competition measure are provided in Table 4.11 in accordance with the time periods: 1980-1994, 1995-2008, and 2000-2008. Other estimates of the square of the competition measure, the incremental R&D stock ratio, and the total IT investment ratio are omitted to turn our focus on competition effects on TFP growth. Table 4.11 indicates that:

1. The competition effects of the manufacturing industries are significantly positive and the estimates are highly robust throughout the sample period, 1980-2008. In addition, the degree of the competition effects becomes slightly larger in 1995-2008 than in 1980-1994.

2. It is not absolutely certain whether competition positively affects TFP growth of the non-manufacturing industries both in 1980-1994 and 1995-2008. (The FD-IV estimates suggest that the competition measure is not significant at the 10% level.) If we suppose that there are competition effects on TFP growth in the non-manufacturing industries, the degree is much smaller (about one tenth) than that of the manufacturing industries. Moreover, notably enough, the competition effects cannot be observable in 2000-2008 in the least.

Whereas the result concerning the manufacturing industries is identical with that derived from FE estimation except for the numerical values of the coefficients (the FD estimates are totally larger than the FE estimates), FD estimation leads to somewhat a different result from the previous analysis for the non-manufacturing industries. More precisely, on the one hand, FE estimation shows that the TFP growth rate may be negatively affected by competition in 1995-2008. On the other hand, although FD estimation indicates a slightly positive competition effect both in 1980-1994 and 1995-2008, the estimates of FD-IV estimation, which are preferred by a series of the tests, are not significant at the 10% level in these periods. Additionally, numerical significance is rather low as compared to the manufacturing industries. In the period of 2000-2008, there is inconclusive evidence that competition has an impact on TFP growth for both FD and FD-IV estimation. Consequently, if we rely on the result of FD (FD-IV) estimation, the conclusion should be rewritten such that the positive or negative competition effects in the non-manufacturing industries do not work strongly in the latter half of the sample period, and in particular, the competition effect is non-observable in 2000-2008 possibly due to the reasons that were mentioned in Subsection 4.5.4.

References

[1] Aghion, P., Bloom N., Blundell, R., Griffith, R., and Howitt, P. (2005), “Competition and Innovation: An Inverted-U Relationship", *Quarterly Journal of Economics*, 120(2), pp. 701-728

[2] Aghion, P., Blundell, R., Griffith, R., Howitt, P., and Prantl, S. (2009), “The Effects of Entry on Incumbent Innovation and Productivity", *The Review of Economics and Statistics*, 91(1), pp. 20-32

[3] Aghion, P. and Griffith, R. (2005), *Competition and Growth: Reconciling Theory and Evidence*, The MIT Press, Cambridge, MA

[4] Ahn, S. (2002), “Competition, Innovation and Productivity Growth: A Review of Theory and Evidence", OECD Economics Department Working Papers, Paper No. 317, OECD, available at:
http://www.oecd-ilibrary.org/economics/competition-innovation-and-productivity-growth_182144868160?crawler=true

[5] Arai, K. (2005), “Competition and Sector Growth" (in Japanese), *Journal of Social Science (Shakaikagaku-Kenkyu)*, Institute of Social Science, University of Tokyo, 56(3)(4), pp. 53-87

[6] Arellano, M. and Bond, S. (1991), “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58(2), pp. 277-297

[7] Arrow, K. J. (1962), “Economic Welfare and the Allocation of Resources for Innovation", Chapter 23 in *The Rate and Direction of Inventive Activity*, ed. Nelson, R. R., Princeton University Press, Princeton, NJ, pp. 609-626

[8] Baily, M. N. and Solow, R. M. (2001), “International Productivity Comparisons Built from the Firm Level", *Journal of Economic Perspectives*, 15(3), pp. 151-172

[9] Baltagi, B. H. (2008), *Econometric Analysis of Panel Data*, 4th edition, Wiley, Hoboken, NJ, pp. 58-59

[10] Blundell, R., Griffith, R., and van Reenen, J. (1999), “Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms", *Review of Economic Studies*, 66(3), pp. 529-554

[11] Cabinet Office of Japan (2006), “The Relationship between Progress of Recent Regulatory Reforms and Productivity" (in Japanese), *Report on Structural Re-*

form Assessment, Report No. 6, Cabinet Office of Japan, available at: http://www5.cao.go.jp/j-j/kozo/2006-12/kozo.html

[12] Cohen, W. M. (1995), “Empirical Studies of Innovative Activity", Chapter 6 in *Handbook of the Economics of Innovation and Technological Change*, ed. Stoneman, P, Wiley, Hoboken, NJ, pp. 182-264

[13] Cohen, W. M. (2010), “Fifty Years of Empirical Studies of Innovative Activity and Performance", Chapter 4 in *Handbook of the Economics of Innovation*, Volume 1, eds. Hall, B. H. and Rosenberg, N., Elsevier, Amsterdam, pp. 129-213

[14] Cohen, W. M. and Klepper, S. (1996), “Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D", *Review of Economics and Statistics*, 78(2), pp. 232-243

[15] Crépon, B., Duget, E., and Mairesse, J. (1998), “Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level", NBER Working Paper Series, Paper No. 6696, The National Bureau of Economic Research, available at: http://www.nber.org/papers/w6696

[16] Dasguputa, P. and Stiglitz, J. (1980), “Industrial Structure and the Nature of Innovative Activity", *Economic Journal*, 90(358), pp. 266-293

[17] Flath, D. (2011), “Industrial Concentration, Price-Cost Margins, and Innovation", *Japan and the World Economy*, 23(2), pp. 129-139

[18] Fukao, K. (2012), *Lost Two Decades and the Japanese Economy: Revealing the Structural Causes and the Driving Force of Revival* (in Japanese), Nikkei Publishing Inc., Tokyo

[19] Fukao, K., Miyagawa, T., Tokui, J., Inui, T., Hamagata, S. (2008), “An Analysis of the Growth Accounting of the Japanese Economy" (in Japanese), Chapter 1 in *Productivity and Japanese Economic Growth*, eds. Fukao, K. and Miyagawa, T., The University of Tokyo Press, Tokyo, pp. 3-47

[20] Geroski, P. A. (1990), “Innovation, Technological Opportunity, and Market Structure", *Oxford Economic Papers*, 42(3), pp. 586-602

[21] Gilbert, R. J. and Newbery, D. M. (1982), “Preempting Patenting and the Persistence of Monopoly", *American Economic Review*, 72(3), pp. 514-526

[22] Green, E. J. and Poter, R. H. (1984), “Noncooperative Collusion under Imperfect Information", *Econometrica*, 52(1), pp. 87-100

[23] Hart, O. D. (1983), “The Market Mechanism as an Incentive Scheme", *Bell Journal of Economics*, 14(2), pp. 366-382

[24] Hayashi, F. and Prescott, E. C. (2002), “The 1990s in Japan: A Lost Decade", *Review of Economic Dynamics*, 5(1), pp. 206-235

[25] Hoshi, T. and Kashyap, A. (2011), “Why Did Japan Stop Growing?", NIRA Study Report, National Institute for Research Advancement, available at: http://www.nira.or.jp/pdf/1002english_report.pdf

[26] Inui, T., Kawakami, A., and Miyagawa, T. (2012), “Market Competition, Differences in Technology, and Productivity Improvement: An Empirical Analysis Based on Japanese Manufacturing Firm Data", *Japan and the World Economy*, 24(3), pp. 197-206

[27] Kim, Y. G., Fukao, K., and Makino, T. (2010), “The Structural Causes of Japan’s Two Lost Decades" (in Japanese), RIETI Policy Discussion Paper Series, Paper No. 10-P-004, Research Institute of Economy, Trade and Industry, available at: http://www.rieti.go.jp/en/publications/act_pdp_jp2010.html

[28] Kwon, H. U., Fukao, K., and Kim, Y. G. (2008), “R&D and Productivity Growth: An Empirical Analysis Based on Firm-Level Data" (in Japanese), Global COE Hi-Stat Discussion Paper Series, Paper No. 3, Institute of Economic Research, Hitotsubashi University, available at: http://gcoe.ier.hit-u.ac.jp/research/discussion/2008/gde08-003.html

[29] Ministry of Economy, Trade and Industry of Japan (2015), *White Paper 2015 on Manufacturing Industries* (in Japanese), Ministry of Economy, Trade and Industry of Japan, Tokyo, available at: http://www.meti.go.jp/english/report/data/gIT2015maine.html

[30] Morikawa, M. (2011), “Economies of Density and Productivity in Service Industries: An Analysis of Personal-Service Industries Based on Establishment-Level Data", *Review of Economics and Statics*, 93(1), pp. 179-192

[31] Morikawa, M. (2014), *The Productivity Analysis in Service Industries* (in Japanese), Nippon-Hyoron-Sha Co., Ltd., Tokyo

[32] Motohashi, K., Funakoshi, M., and Tohei, A. (2005), “The Quantitative Analysis of Competition, Innovation, and Productivity" (in Japanese), Joint Research Report, Competition Policy Research Center, Japan Fair Trade Commission, available at: http://www.jftc.go.jp/cprc/reports/index.files/cr0205.pdf

[33] Nakanishi, Y. and Inui, T. (2007), “Deregulation and Productivity in Japanese Industries" (in Japanese), CEI Working Paper Series, Paper No. 2007-3, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University, available at: http://cei.ier.hit-u.ac.jp/Japanese/pdf/wp_2007/wp2007-3.pdf

[34] Nickell, S. J. (1996), “Competition and Corporate Performances", *Journal of Political Economy*, 104(4), pp. 724-746

[35] Nishimura, K. G. and Shirai, M. (2003), “Can Information and Communication technology Solve Japan’s Productivity Slowdown Problem?", *Asian Economic Papers*, 2(1), pp. 85-136

[36] Okada, Y. (2005), “Competition and Productivity in Japanese Manufacturing Industries", *Journal of the Japanese and International Economics*, 19(4), pp. 586-616

[37] Porter, M. E. and Sakakibara, M. (2004), “Competition in Japan", *Journal of Economic Perspectives*, 18(1), pp. 27-50

[38] Rotemberg, J. J. and Saloner, G. (1986), “A Super Game-Theoretic Model of Price Wars during Booms", *American Economic Review*, 76(3). pp. 390-407

[39] Scharfstein, D. (1988), “Product Market Competition and Managerial Slack", *Rand Journal of Economics*, 19(1), pp. 147-155

[40] Scherer, F. M. (1965), “Firm Size, Market Structure, Opportunity, and the Output of Patented Innovations", *American Economic Review*, 55(5), pp. 1097-1125

[41] Scherer, F. M. and Ross, D. (1990), *Industrial Market Structure and Economic Performance*, 3rd edition, Rand Mcnally, Boston, MA, pp. 652

[42] Schmidt, K. M. (1997), “Managerial Incentives and Product Market Competition", *Review of Economic Studies*, 64(2), pp. 191-213

[43] Schumpeter, J. A. (1942), *Capitalism, Socialism, and Democracy*, Routledge, London

[44] Symeonidis, G. (1996), “Innovation, Firm Size, and Market Structure: Schumpeterian Hypotheses and Some New Themes", *OECD Economic Studies*, 27(2), pp. 35-70

[45] Wei, Y. and Liu, X. (2006), “Productivity Spillovers from R&D, Exports and FDI in China’s Manufacturing Sector", *Journal of International Business Studies*, 37(4), pp. 544-557

[46] Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, The MIT Press, Cambridge, MA, pp. 321-326

[47] Yagi, M. and Managi, S. (2013), “Competition and Innovation: An Inverted-U Relationship Using Japanese Industry Data", RIETI Discussion Paper Series, Paper No. 13-E-062, Research Institute of Economy, Trade and Industry, available at: http://www.rieti.go.jp/en/publications/summary/13070005.html

[48] Yoshino, N. and Sakakibara, E. (2002), “The Current State of the Japanese Economy and Remedies", *Asian Economic Papers*, 1(2), pp. 110-126