

# **Three Essays on R&D Investment**

**Massoud Khazabi**

Thesis submitted to the  
Faculty of Graduate and Postdoctoral Studies  
in partial fulfillment of the requirements  
for the Ph.D. degree in Economics

Department of Economics  
Faculty of Social Sciences  
University of Ottawa

© Massoud Khazabi, Ottawa, Canada, 2011

## **Acknowledgements**

I am deeply indebted to my supervisor, Professor N. V. Quyen for his inspiring support, advice and patience during the preparation of this dissertation. This never would have been accomplished without his help and support.

I would also like to express my appreciation to the evaluation committee members, Professor G. Atallah, Professor P. Makdissi, Professor M. Voia, and Professor D. Sen, the staff and faculty members of the Departments of Economics at the University of Ottawa and Carleton University for their advice, help, and support.

*To my beloved wife, Leila, for her support;*

*To my children, the spirits of my life, Kian and Kobra, for their encouraging  
smiles;*

*To the souls of my best friends and best teachers, my father and my brother,  
Hassan and Mehrdad, for all the good memories;*

*To the best mom in the world, Shirin, for her devotions; and,*

*To my brother, Keyvan, in hope of his return.*

## Table of Contents

|                                                                                               |             |
|-----------------------------------------------------------------------------------------------|-------------|
| <b>ABSTRACT .....</b>                                                                         | <b>VI</b>   |
| <b>INTRODUCTION.....</b>                                                                      | <b>VIII</b> |
| REFERENCES .....                                                                              | XIII        |
| <br>                                                                                          |             |
| <b>1 FUNDAMENTAL SOURCES OF LONG-RUN LABOUR PRODUCTIVITY<br/>IMPROVEMENTS IN CANADA .....</b> | <b>1</b>    |
| 1.1 INTRODUCTION.....                                                                         | 1           |
| 1.2 TRENDS OF LABOUR PRODUCTIVITY IN CANADA.....                                              | 5           |
| 1.3 THE DATA AND METHODOLOGY.....                                                             | 6           |
| 1.4 THE RESULTS .....                                                                         | 10          |
| 1.5 CONCLUSION .....                                                                          | 15          |
| ANNEX A .....                                                                                 | 17          |
| ANNEX B: TEST RESULTS.....                                                                    | 19          |
| REFERENCES .....                                                                              | 27          |
| <br>                                                                                          |             |
| <b>2 R&amp;D SPILLOVERS, INNOVATION, AND ENTRY .....</b>                                      | <b>31</b>   |
| 2.1 INTRODUCTION.....                                                                         | 31          |
| 2.2 THE GENERAL FEATURES OF THE MODEL .....                                                   | 37          |
| 2.3 THE POST-INNOVATION EQUILIBRIUM.....                                                      | 39          |
| 2.4 THE EQUILIBRIUM IN THE INNOVATION STAGE.....                                              | 42          |
| 2.5 ENTRY .....                                                                               | 48          |
| 2.6 THE PROPERTIES OF THE EQUILIBRIUM .....                                                   | 49          |
| 2.7 A NUMERICAL SIMULATION .....                                                              | 55          |
| 2.8 CONCLUSION .....                                                                          | 61          |
| ANNEX A: THE PROOF OF LEMMA 3.....                                                            | 61          |
| ANNEX B: THE PROOF OF PROPOSITION 4 .....                                                     | 62          |
| ANNEX C: THE PROOF OF LEMMA 4.....                                                            | 64          |
| REFERENCES .....                                                                              | 68          |

|          |                                                                                           |           |
|----------|-------------------------------------------------------------------------------------------|-----------|
| <b>3</b> | <b>THE SEARCH FOR NEW DRUGS: A THEORY OF R&amp;D IN THE PHARMACEUTICAL INDUSTRY .....</b> | <b>70</b> |
| 3.1      | INTRODUCTION .....                                                                        | 70        |
| 3.2      | THE MODEL.....                                                                            | 79        |
| 3.2.1    | <i>The R&amp;D Process</i> .....                                                          | 79        |
| 3.2.2    | <i>Patent Breadth</i> .....                                                               | 85        |
| 3.2.3    | <i>Preferences</i> .....                                                                  | 87        |
| 3.3      | THE PRICES OF DRUGS.....                                                                  | 89        |
| 3.4      | R&D AND INFORMATION EXTERNALITIES.....                                                    | 94        |
| 3.4.1    | <i>One Remaining Unscreened Chemical Compound</i> .....                                   | 96        |
| 3.4.2    | <i>Two Remaining Unscreened Chemical Compounds</i> .....                                  | 102       |
| 3.4.3    | <i>Three or More Remaining Unscreened Chemical Compounds</i> .....                        | 118       |
| 3.5      | CONCLUDING REMARKS .....                                                                  | 119       |
|          | ANNEX A: THE PROOF OF PROPOSITION 1 .....                                                 | 121       |
|          | ANNEX B: THE PROOF OF PROPOSITION 2 .....                                                 | 123       |
|          | ANNEX C: THE PROOF OF PROPOSITION 3 .....                                                 | 124       |
|          | ANNEX D: THE PROOF OF LEMMA 2.....                                                        | 124       |
|          | ANNEX E: THE PROOF OF PROPOSITION 4 .....                                                 | 125       |
|          | ANNEX F: THE PROOF OF LEMMA 3 .....                                                       | 126       |
|          | ANNEX G: THE PROOF OF PROPOSITION 5 .....                                                 | 128       |
|          | ANNEX H: THE PROOF OF PROPOSITION 6.....                                                  | 129       |
|          | ANNEX I: THE PROOF OF PROPOSITION 7 .....                                                 | 131       |
|          | ANNEX J: THE PROOF OF PROPOSITION 8 .....                                                 | 133       |
|          | REFERENCES .....                                                                          | 135       |

## **Abstract**

This dissertation consists of three essays on Research and Development investment. The first essay titled “Fundamental Sources of Long-run Labour Productivity Improvements in Canada” examines the importance of Research and Development activities, as well as the stock of public infrastructure, and economic openness as sources of growth in labour productivity in the Canadian economy within the last four decades.

Time series data are used to estimate an econometric model to capture the interaction among labour productivity, R&D investments, public infrastructure, and trade openness. Extensive tests of data quality, choice of model, and stability of the model are performed to enrich the findings. The results suggest that the type of capital involved has a significant impact on the extent of labour productivity and growth. Innovation, the major characteristic of a knowledge-based economy, improves labour productivity both in the short and long run. The effect of trade openness is also discussed in depth.

The second paper titled “R&D Spillovers, Innovation, and Entry” extends a theoretical framework to analyze the impact of R&D spillovers on entry and the resulting equilibrium market structure. It is shown that the degree of spillovers plays a fundamental role on the number of firms entering the market, their R&D activities, and social welfare. The analysis suggests that social welfare is maximized at some intermediate degree of spillovers. The policy implication of this result is that neither complete protection of intellectual property right nor lax enforcement of patent laws is socially optimal.

The third paper titled “The Search for New Drugs: A Theory of R&D in the Pharmaceutical Industry” uses a dynamic model of optimal patent design and in the presence of information externalities studies the evolution of technological progress

in the context of a pharmaceutical industry. The preceding literature on the topic works with only one brand, the brand with the highest quality. As well, the demand is assumed to be completely inelastic. In the conventional models of patent design the role of competitive fringe firms is also discussed implicitly. The model presented in this research is a continuous in-time dynamic model which provides a rigorous structure for studying the context. It considers several differentiated products, both those whose patents are still in force and those whose patents have already expired, at any point in time. Furthermore the demand for a brand is taken to be a function of income, its price, and the prices of other brands. The interaction of the fringe firm with other patent-holding firms is also explicitly considered under this framework. Unlike the previous literature on the context, the model incorporates both product and process innovation concepts and provides real guidelines to measure the patent breadth. Under this structure, pharmaceutical firms with an active drug discovery program behave strategically in their R&D and in the product markets.

## **Introduction**

The Oxford dictionary defines innovation as a new method, idea, and product, and Research and Development (as its synonym) as work directed towards innovation in, and improvement of, products and processes. The economic literature differentiates between two types of innovation activities – process innovation vs. product innovation. Process innovation particularly focuses on the process of production and improvement of equipment and plants by which firms can gain productivity and improve the quality and reliability of their products. This, sometimes, may facilitate product innovation by advancing the firms' capacities. Product innovation, however, tends to narrow its sight to the design, development and establishment of new products. In this sense, the focus of the second chapter of this dissertation is on process innovation while those of the first and the last chapters are on both process and product innovations.

The first essay of this dissertation defines public infrastructure and then uses the definition to tabulate the time series to represent the variable. Observations on the variable along with those for labour productivity, economic openness, R&D expenditure, and total hours worked over a forty-three-year period are then used to estimate an econometric model and examine the sources of labour productivity growth in the Canadian economy. In the short run all three major explanatory variables – the stock of public infrastructure, economic openness and R&D expenditure – show positive effects on labour productivity. The coefficients of the first and second lagged values of trade openness, however, are indicative of a negative relationship between mid-term economic openness and labour productivity. A negative correction factor confirms an adjustment period of approximately four years for any deviation from the equilibrium status of labour productivity.

The strong relationship between labour productivity and the stock of public infrastructure suggests that the latter is the main source of long-run productivity

growth. R&D costs and economic openness were also found to have a positive long-run relationship with labour productivity.

The findings of the paper are hence along the lines of neoclassical growth theory insofar as accumulation of infrastructure investment explains the majority of growth in labour productivity (see Solow 1956 & 1957 for more detail). The remainder of the productivity growth in the Canadian context is explained by technological progress resulting from R&D expenditures as well as economic openness, as suggested by the new growth theory (see Romer, 1986, 1987 & 1990, Lucas, 1988 and Grossman and Helpman, 1991 for more detail).

The second paper focuses on the R&D spillovers and process innovation literature. In the literature efforts are mostly focused on the comparative R&D expenditures and the relative social welfare between non-cooperative and cooperative R&D. The question of how innovation is affected when there is more competition, i.e., when the number of firms rises, is ignored by most researchers, except for De Bondt *et al.* (1992), who discussed how the number of firms affects innovation. However, the question of how R&D spillovers affect entry was not addressed by these researchers. The second chapter attempts to fill part of this lacuna by endogenizing the number of firms. More specifically, the model addresses the following questions. First, how does the degree of spillovers affect the equilibrium number of firms? Second, how does the degree of spillovers affect the equilibrium market structure? Third, when is the equilibrium symmetric and when is it not symmetric, and in the case of asymmetric equilibria, how many firms choose to incur a positive amount of own R&D cost and how many firms choose to free ride on the R&D activities of others?

The model analyzes the influence of competition on innovation in a two-stage game played by a number of firms producing a homogeneous good. In the first stage of the game, the firms carry out R&D activities to lower their production cost. It is assumed that before innovation, all the firms have the same marginal cost. In the second stage,

the firms compete in the product market according to the Cournot model of competition. All the firms act non-cooperatively in both stages of the game. In the R&D stage, each firm runs its own research lab, and takes into account the natural spillovers that flow among firms in a strategic manner. It is then shown that if the degree of spillovers is low, only a finite number of firms enter the market, and after entry – all the firms expend the same amount of their own resources on R&D. This type of equilibrium is referred to as a symmetric equilibrium with innovation throughout the paper. The intuition behind this result is not hard to understand. When the degree of spillovers is low, a firm cannot rely on the R&D externalities generated by the other firms to lower its own marginal cost. If the degree of spillovers is high, an infinite number of firms enter the market, and after entry none of the firms chooses to expend any of its own resource on R&D. The equilibrium is a perfectly competitive equilibrium without innovation. For intermediate values of the degree of spillovers rises, the equilibrium is an asymmetric equilibrium under which some firms choose to expend their own resources on R&D, while others choose not to do so. Furthermore, all the firms that choose to expend their own resources on R&D choose the same amount of own expenditure on R&D. In the literature on R&D spillovers and innovation, the number of firms in the market is taken as exogenous, and the equilibrium is presumed to be symmetric. In our model, the number of firms that enter the market is endogenous, and varies according to the degree of spillovers. The endogenization of the number of firms yields different types of equilibrium market structures – symmetric equilibrium with innovation, asymmetric equilibrium with innovation, and perfect competition without innovation – and these are novel results in the field of R&D spillovers and innovation.

The welfare analysis presented in this paper suggests that social welfare rises with the degree of spillovers when it is low, reaches a maximum when the degree of spillovers enters its intermediate range, and then declines to the level associated with the competitive equilibrium without innovation. The policy implication of this result is that the intellectual property right should not be fully protected and the enforcement

of patent laws should not be too lax. The optimal degree of protection should reflect the right trade-off between allocative and dynamic efficiency.

The third essay develops a model of optimal patent design to study the evolution of technological progress in the context of a pharmaceutical industry. In the earlier literature on patent design, all efforts were concentrated on finding the optimal length and breadth of a patent. The role of information externalities was not considered. Furthermore, the analysis was carried out in the context of a single innovation, and the inherent uncertainties in R&D activities were ignored. An exception was Scotchmer and Green (1990), who formulated a simple dynamic economic model that incorporates patent breadth and the incentive of firms to keep secret their R&D results. O'Donoghue et al. (1998) also formulated a model in which firms sequentially improve the products of each other through time, and technological progress occurs when a non-infringing innovation displaces a patented product.

The purported goal of a patent policy is to encourage technological progress. Yet, its impact on the pace of technological progress has largely been ignored by researchers in this field. Compared to the models of R&D in the literature on trade and technological progress, the model of R&D found in the literature of optimal patent design is much less sophisticated. In the research program on trade and technological progress of Eaton and Kortum (2001), ideas (discoveries) arrive in time according to a Poisson process whose parameter depends on the stock of knowledge – taken to be the cumulative effort of research workers. However, this aggregate variable is incapable of capturing the impact of information externalities. To formalize the information externalities in a rigorous manner, we need to descend to a more disaggregate level and model how the findings from the R&D activities of one research organization can have a positive impact on the R&D activities of another research organization. To tease out the subject, we have chosen to model R&D activities as the process of searching for a drug to treat a disease. The main reason for formulating the R&D process in the context of the pharmaceutical industry is because

much knowledge has been gained in the process of drug discovery, and this allows for a less abstract modelling of the R&D process.

In conventional models of patent design, only one brand – the brand with the highest quality – is available on the market at any time. Furthermore, demand is completely inelastic: each consumer, regardless of income and regardless of the price of the brand offered on the market, buys exactly one unit of the brand. In the model we formulate, several differentiated products – the products whose patents are still in force and the products whose patents have expired – are available at any time on the market. Furthermore, the demand for a brand depends on income, its own price, and the prices of the other brands. In our model, the pharmaceutical firms with an active drug discovery program behave strategically in both R&D and in the product market. The model also considers a competitive fringe, which manufactures the brands whose patents have expired. The analysis of R&D is then represented under this framework.

## References

- De Bondt, R., P. Slaets, and B. Cassiman (1992)**, “The Degree of Spillovers and the Number of Rivals for Maximum Effective R&D,” *International Journal of Industrial Organization*, 10, 35-54.
- Eaton, J. and S. Khortum (2001)**, “Technology, Trade, and Growth: A Unified Framework,” *European Economic Review*, 45, 742-755.
- Grossman, G. M. and E. Helpman (1991)**, “Innovation and Growth in the Global Economy,” *MIT Press*, Cambridge.
- Khazabi, M. (2008)**, “On R&D investment,” *IUP Journal of Managerial Economics*, 6(2), 54-68.
- Lucas, R. E. (1988)**, “On the Mechanics of Development Planning,” *Journal of Monetary Economics*, 22, 3-42.
- O’Donoghue, T., S. Scotchmer, and J. Thisse (1998)**, “Patent Breadth, Patent Life, and the Pace of Technological Progress,” *Journal of Economics & Management Strategy*, 7(1), 1-32.
- Romer, P. (1986)**, “Increasing returns and long – run growth,” *The Journal of Political Economy*, 94(5), 1002-1037.
- **(1987)**, "Growth Based on Increasing Returns Due to Specialization," *American Economic Review*, 77(2), 56-62
- **(1990)**, "Are Nonconvexities Important for Understanding Growth?," *American Economic Review*, 80(2), 97-103
- Schumpeter, J. A. (1942)**, “Capitalism, Socialism and Democracy,” New York, Harper & Row, 381 pp.
- Scotchmer, S. And J. Green (1990)**, “Novelty and Disclosure in Patent Law,” *The Rand Journal of Economics*, 21, 131-146.
- Solow, R. M. (1956)**, “A Contribution to the Theory of Economic Growth,” *Quarterly Journal of Economics*, 70, 65-94.
- **(1957)**, “Technical Change and the Aggregate Production Function,” *The Review of Economics and Statistics*, 39(3), 312-320

# **1 Fundamental Sources of Long-run Labour Productivity Improvements in Canada**

## **1.1 Introduction**

The OECD manual defines labour productivity as “the ratio of volume measure of output to a volume measure of input,” for which the volume measure of output is the constant price gross domestic product and the volume measure of input is usually the total hours worked.

In economic literature, labour productivity has a root to the origin of the growth theory as it has always been considered a major contributor to living standards and economic growth (see Harrod, 1939 and Domar, 1946). In his pioneering work, Solow (1956) delineated labour productivity as a function of capital, and argued that capital deepening (increase in accumulated capital per worker) could improve labour productivity and growth. Since the seminal work of Solow, *op. cit.*, the debate on productivity has taken many different directions. Shell (1967) motivated the endogenous growth literature by adding R&D – as a publicly-funded activity – to the Solow-Swan growth model, and hence opened the way for future research, such as Romer (1986 & 1990), Coe and Helpman (1995), and Grossman and Helpman (1999). Under the new growth paradigm, Coe and Helpman (1995) relate cross-country R&D spillovers to productivity growth and open a new era of thinking on trade and growth. Their seminal work was then followed by Keller (1998), Frankel and Romer (1999), and Dowrick and Golley (2004).

The pioneering work of David Aschauer in 1989, in which infrastructure investment was introduced into the context as were thoughts about the importance of human

capital in improving the productivity (Lucas, 1988), also added new research areas to the literature. Aschauer (1989) studies the relationship between public investment on infrastructure and total factor productivity. Using a Cobb-Douglas production function and US time series for the period of 1949-1985, he concludes that a 1% increase in non-military public capital has a 0.36% to 0.49% impact on the total factor productivity. DeLong and Summers (1991), Levine and Renelt (1992), Clarida (1993), Lynde and Richmond (1993), Gramlich (1994), Blomstrom *et al* (1996), and Otto and Voss (1996, 1998, 2003) test similar hypotheses to confirm the importance of fixed capital and public infrastructure<sup>1</sup> as major drivers of productivity and growth.

In contrast to the scope of international work published on the context, few Canadian studies try to capture the impact of infrastructure on productivity growth. Those who emphasize its importance face major limitations to their research, as the availability and quality of infrastructure-related databases in Canada is seriously limited. These limitations are mostly attributable to a lack of precise definitions for infrastructure. Many scholars such as Wylie (1996), Harchaoui & Tarkhani (2003), Bin Dong (2005), Satya *et al* (2004), and Brox *et al* (2005) study the impact of public capital/public infrastructure on productivity growth, without necessarily providing a rigorous definition for the variable(s) or, in some cases, without even making a distinction between the two concepts. Examining the literature just described, it is easy to assert confusion, and contradiction around their understanding of the term “public infrastructure” and its importance. While it is recognized that public infrastructure should include not only physical capital, but also human, and social capital, the literature provides little guidance on how to conceive of these inclusions. Most remarkably, while everyone seems to know what public infrastructure is, precise definitions are in short supply.

---

<sup>1</sup> The term “public infrastructure” is used to underline the public sector special interest on the variable rather than to specify its ownership. A more precise definition is provided on page 7, which serves to clarify the matter more rigorously.

The problem is partly due to the fact that traditional definitions of public infrastructure are restricted to *publicly owned* assets. Defining infrastructure in this way therefore, ignores the current trends towards privatization and public-private partnerships. As there is a serious shortfall in providing an accurate measurement which goes further than roads, rails, and cables, the real effect of infrastructure on economic growth and prosperity has always been underestimated and/or neglected. Stiroh (2000) considers this as a major caveat of the research on the context and suggests a broader definition of the concept as the appropriate solution.

The literature is also murky when it comes to the categories within public infrastructure which are beneficial to productivity growth. This work deals with these lacunae and seeks to provide policy guidelines to help prevent a failure to realize investment in public infrastructure, which is critical to growth and prosperity in the Canadian economy. The major contribution of this paper to the existing literature is a concrete, comprehensive, and precise working definition for public infrastructure, which is then used to generate the stock of public infrastructure time series. In this sense, the paper is among a few articles in the field which aim to provide a rigorous working definition for the variable. To the authors' best knowledge the only other subsequent work published on the topic is that of Torrasi in 2009. Torrasi (2009) emphasizes the importance of having a rigorous definition for public infrastructure and briefly discusses the measurement issues related to tabulating the time series.

The study also examines the sources of Canadian labour productivity growth between 1961 and 2003. Time series data are used to develop an econometric model that captures the interaction between labour productivity and R&D, as well as the roles for the stock of infrastructure and trade openness. The results suggest that the type of capital involved has a significant effect on the extent of labour productivity and its rate of growth. Labour productivity is improved by innovation, which is the major characteristic of a knowledge-based economy, in both the short and long terms. Trade

openness, meanwhile, has a positive effect on labour productivity in both the short and long terms, but not in the medium term.

The paper also introduces the use of the Autoregressive Distributed Lag (ARDL) approach to cointegration to the context as its other major contribution to the literature. The robustness of the results is then confirmed using the Dynamic Ordinary Least Squares (DOLS) approach. A potential weakness of all previously-used techniques in the field is that they required the time series to be of the same order of integration. Those studies, which recognized this requirement, used unit root tests to support the presence of a unit root among the series, however the low power of such tests (especially in satisfying the persistency of the series) has long been an area of concern. The inferences made, therefore, condition the hypothesis testing on a low power, and, consequently, are on rather shaky technical ground. As a radical statistical innovation, the ARDL approach to cointegration does not suffer from this shortcoming. It requires no assumption or investigation on the order of integration of the series, and thus bases the discussion on a stable statistical ground. As this is the case in some of the previous studies (i.e. Madden *et al* 1998, and Wylie 1996), failure to apply this method may have resulted in spurious results. Both the model and the approach are flexible enough to take into account the complexity of the relationship. The ARDL approach to cointegration also deals with many technical issues raised in the previous works done in the same context by Canadian scholars (e.g. the problem of multicollinearity in Wylie, 1996). This work hence provides an exclusive basis of comparison for the existing European and Australian literature.

The paper is organized as follows: Section 2 provides a trend analysis of labour productivity in Canada within the period of study. Section 3 presents the details on the data and methodology used in this paper. The results of the paper are presented in Section 4, and Section 5 concludes.

## 1.2 Trends of Labour Productivity in Canada

Labour productivity measures the efficiency with which labour is employed in the production process, and growth in labour productivity is the difference between the growth in output and the growth in labour input. Labour productivity in the Canadian economy rose by an average of 2% annually over a period of 43 years – between 1961 and 2003 – which means that the output per worker doubled over a 35-year period. The 2% annual growth in labour productivity during this period constitutes approximately 50% of GDP growth.

Figure 1 shows the evolution of labour productivity in Canada between 1961 and 2003. As can be seen from the figure, labour productivity rose smoothly from 1961 to 1978. It fell slightly in 1979 and 1980 before improving again in 1981. This pattern was repeated in 1986 and 1996, but was followed by immense improvements in 1987 and 1997.

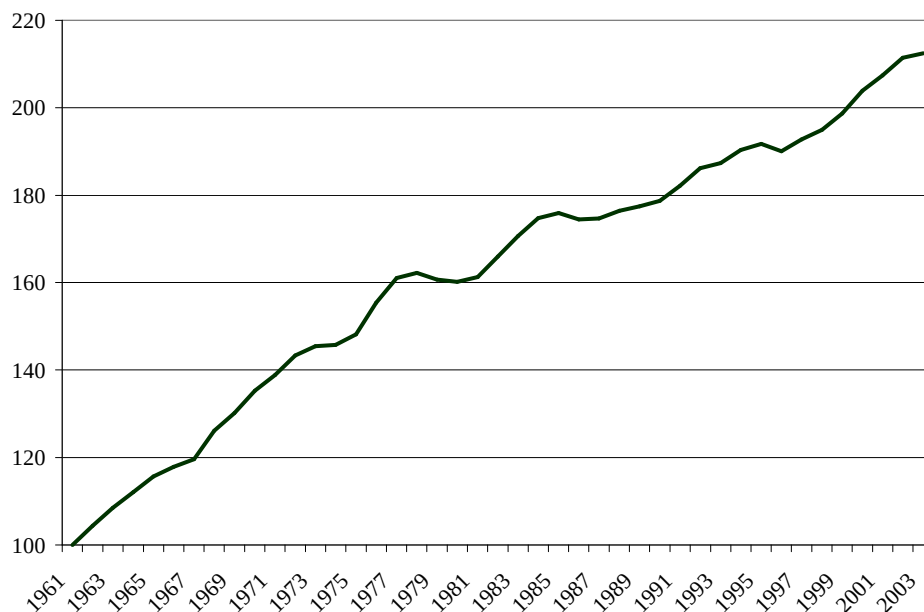


**Figure 1. Trend of labour productivity in the Canadian economy (1961-2003)**

In addition to public capital/public infrastructure, growth in Canadian labour productivity is sometimes explained by stock of machinery and equipment, engineering construction, and building construction (see Macklem, 2003 and Abdi, 2008, for example). Technological and management advances resulting from innovation activities have also been recognized as sources of improvement in labour productivity (see Gera *et al*, 1999, for example). Improvement in trade is responsible for economies of scale; a more competitive market for the goods and services results in a more efficient production process, and therefore increases labour productivity (see Trefler, 2004 and Lilveeva, 2008, for example).

### 1.3 The Data and Methodology

Using the production function approach of Aschauer (1989), the evolution of labour productivity is broken down in the following manner:

$$(1) \quad \frac{Y}{L} = f\left(\frac{T}{L}, \frac{IE}{L}, \frac{RD}{L}\right),$$

where  $Y/L$  is labour productivity,  $T/L$  is public infrastructure per hour worked,  $IE/L$  is economic openness per hour worked, and  $RD/L$  is the innovation cost per hour worked.

The Statistics Canada publications are the main source of data in this research. The labour productivity ( $Y/L$ ) time series is from KLEMS (Capital, Labour, Energy, Material & Services) database. The annual data are reported at 1997 constant prices. The annual data on R&D expenditures (RD), imports and exports (IE), and public infrastructure (T) are tabulated from CANSIM II (Canadian Socio-economic Information Management System) database and are reported at 1997 constant prices. The period of study is 1961 to 2003. The restriction is being imposed by data availability. All the variables are measured in natural logarithms.

As a working definition, public infrastructure is defined as the part of stock of capital in the economy that facilitates the fundamental social and economic activities in that economy. Based on which aspect of the productivity growth it impacts, public infrastructure is classified into 8 major categories: *Transportation Infrastructure* (e.g. airports, rail tracks, roads and highways, marine ports, and associated facilities); *Energy Infrastructure* (e.g. energy and power generations, and transmission and distribution networks); *Telecommunication Infrastructure* (e.g. transmission and switching facilities, as well as wired and wireless networks); *Water and Sewer Supply Infrastructure* (e.g. distribution networks, and treatment facilities); *Health Infrastructure* (e.g. hospitals, clinics, health research facilities, and laboratories); *Educational Infrastructure* (e.g. schools, colleges, universities, and libraries); *Recreational Infrastructure* (e.g. parks, museums, and sport and other recreational facilities); and *Public Safety/Security Infrastructure* (e.g. fire stations, police facilities, courts, correctional facilities, and national defence facilities.)

Innovation costs/investments are the expenses which refer to discovering new knowledge about products, processes, and services. This knowledge would be applied to create new and improved products, and/or processes that fill market needs. Most international standards ask firms to either expense their development costs immediately or capitalize them if they meet certain criteria. Canadian standards require TSX-listed firms to expense their development costs. As an exception, firms that would like to defer their innovation expenditures have to meet some provisional norms such as clarity of market definition, resource capability for completing their R&D projects, technical achievability of R&D outcome, and specified amortization period. Canada is evaluating a plan in which complete capitalization of innovation costs is authorized. For a more detailed discussion on this topic, the reader can consult Khazabi (2008).

Economic openness refers to the economy's contribution to international trade. The proxy variable used in this work (similar to that used by OECD) is the ratio of (imports + exports) over GDP, or simply the proportion of total trade in GDP.<sup>2</sup>

The variables of interest first have to be tested for stationarity. ADF, Philip-Perron, Dickey-Fuller GLS and Elliott-Rothenberg-Stock Point Optimal tests of unit root can satisfy this purpose. Classical methods of estimation (e.g. traditional Autoregressive Distributed Lag (ARDL) method) are only recommended if the stationarity of variables are satisfied. Under the condition, estimation and inference about the long-run properties of the model will be performed using standard asymptotic normal theory. In presence of a unit root, however, the analysis becomes more complicated, as an alternative estimation method would be necessary. This has been the focus of more recent theoretical literature on cointegration or analyzing the long-run relationship between  $I(1)$  variables following a first-order integration process (see Engle-Granger's residual based test, 1987, Johansen's maximum likelihood-based test, 1988, and the Johansen-Juselius trace test, 1990). All of these cointegration procedures however require that regressors to have the same order of integration (i.e.  $I(1)$  or  $I(2)$  process). Furthermore, these methods do not include the information on structural breaks in the data, and generally suffer from a low power. The Engle-Granger's method of cointegration is also known to be inefficient when there are more than two variables under consideration (H. Pesaran and B. Pesaran, 1997).

These problems once again draw attention to the use of the traditional ARDL approach for the analysis of long-run relations in presence of a unit root. In a series of articles by Pesaran and Shin (1998), Pesaran and Pesaran (1999), and Pesaran *et al* (2001) this task was carried out. These researchers showed that augmented OLS-based ARDL estimators were asymptotically consistent, and consequently the asymptotic theory was applicable to make inferences about the long-run use of ARDL results. The advantage of the ARDL approach is that it can be applied to any group of

---

<sup>2</sup> OECD ANA database

regressors, regardless of their integration order, whether the variable is an I(0), I(1) or even fractionally integrated (Pesaran and Shin, 1998). Moreover, with some transformation, a dynamic error correction model (ECM) can be derived easily from the ARDL approach to cointegration (Annex A). As discussed by Engle and Granger (1987), the ECM puts the short-run and long-run results together, and no information is lost during the estimation process as a result.

The ARDL procedure starts with a cointegration test (the bound test of no cointegration). If the null hypothesis is rejected at the selected critical level, the ARDL method estimates  $(p+1)^k$  different regressions in order to find the optimal lag length of each variable, with  $p$  and  $k$  being the maximum number of lag lengths and number of the variables, respectively. The best model would then be determined using information criteria such as Akaike's information criteria (AIC) and the Schwartz-Bayesian Criteria (SIC). SIC selects the parsimonious model, which is the model with the smallest possible lag length. The long-run relationship between the variables is subsequently estimated, based on the selected ARDL model. Estimation of the error correction model is the next step. The ECM representation of the variables provides both the short-run and the adjustment coefficient to the long-run equilibrium. The adjustment coefficient, or correction coefficient, refers to the speed of adjustment between status quo and the new long-run equilibrium caused by a short-run shock.

Specific to this paper is the intention to examine the relationship between labour productivity and three explanatory variables: the stock of public infrastructure, R&D expenditure, and economic openness. Hence, the ECM representation of the model could be written as

$$(2) \quad \Delta \ln YL_t = \alpha + \sum_{i=1}^{n-1} \mu_i \Delta \ln YL_{t-i} + \sum_{i=0}^{m-1} \eta_i^1 \Delta \ln TL_{t-i} + \sum_{i=0}^{m-1} \eta_i^2 \Delta \ln RDL_{t-i} + \sum_{i=0}^{m-1} \eta_i^3 \Delta \ln IEL_{t-i} \\ + \tau \left( \ln YL_{t-1} + \hat{\theta}_0 + \hat{\theta}_1^1 \ln TL_{t-1} + \hat{\theta}_1^2 \ln RDL_{t-1} + \hat{\theta}_1^3 \ln IEL_{t-1} \right) + \varepsilon_t$$

In equation (2), the terms with the summation signs correspond to the error correction dynamic of the model while the terms inside the parenthesis represent the long-run relationship between labour productivity and its components. The term  $\tau$  is the adjustment coefficient, or error correction coefficient, which refers to the period required for any deviation from the equilibrium status of labour productivity to adjust.

The extent of the fit of the ARDL model can be checked by conducting diagnostic and stability tests. The diagnostic tests are appropriate for testing serial correlation, functional form, normality, and heteroscedasticity. The cumulative sum of recursive residuals (CUSUM) and the cumulative sum of squares of recursive residuals (CUSUMSQ) can be applied to check the stability of the coefficients of the model.

#### 1.4 The Results

ADF, Dickey-Fuller GLS and Elliott-Rothenberg-Stock Point Optimal tests of unit root are carried out on the data. All tests show that variables are non-stationary of order one. This requires using a cointegration approach to test the existence of a long-run relationship between the variables and, thus, the ARDL approach to cointegration is adopted. The bounds test of no cointegration (i.e. the null of  $\hat{\theta}_1^1 = \hat{\theta}_1^2 = \hat{\theta}_1^3 = 0$  which indicates no long-run relationship between labour productivity and its drivers) is then performed. The test statistic can be compared to two asymptotic critical values. If the statistic is larger (smaller) than the upper (lower) critical value, the null hypothesis of no long-run relationship is rejected. This conclusion is made regardless of the underlying order of integration of the variables. In the case that the statistic falls between the two bounds, the result is inconclusive and the objective can be pursued in a second best world when a traditional cointegration approach is used.

The bounds test results in the rejection of the null hypothesis of no cointegration. At the next step the ARDL method estimates 256 different regressions in order to find the optimal lag length of each variable. The best model is then determined using the Schwartz-Bayesian criteria (SIC) and by adopting the Box-Jenkins' model selection approach and employing the principal of parsimony: it is ARDL(1,0,0,3).<sup>3</sup> The sensitivity of the selection is checked by estimating the selected models identified by AIC, which is ARDL(2,0,1,3), and Hannan-Quinn criteria (HQ), which is also ARDL(2,0,1,3). As the coefficients of the lagged values of  $\ln YL$  and  $\ln RDL$  are insignificant in both models, these models are reduced to that suggested by SIC.<sup>4</sup> Table 1 shows the estimated long-run coefficients of the model using the ARDL approach to cointegration.

**Table 1.**  
**Estimated long-run coefficients using the ARDL(1,0,0,3)**

| <i>Regressor</i> | <i>Coefficient</i> | <i>Standard Error</i> |
|------------------|--------------------|-----------------------|
| $\ln TL$         | 0.462***           | 0.116                 |
| $\ln RDL$        | 0.251***           | 0.050                 |
| $\ln IEL$        | 0.208**            | 0.087                 |
| $C$              | 0.408              | 0.420                 |

Note 1: Dependent variable is  $\ln YL$

Note 2: \*\*\*, \*\*, and \* denote significance at 1%, 5%, and 10%, respectively.

Except for the constant, all of the coefficients are significant at the 5% level, and are all inelastic. In the long run, the public infrastructure has the greatest impact. Its estimated coefficient is 0.46, which means that a hypothetical increase of 1% in the stock of public infrastructure per hour worked will increase the productivity of labour

<sup>3</sup> It is also shown that the forecasting accuracy of SIC is higher than AIC (Koehler *et al*, 1998).

<sup>4</sup> The results of the autocorrelation test also confirm the adequacy of the selected model.

input by 0.46% in the long run. The coefficients of  $\ln RDL$  and  $\ln IEL$  are also significantly positive.

Table 2 summarizes the error correction representation of the model as shown in equation (2).

**Table 2.**

**Error Correction representation for the selected ARDL model**

| <i>Regressor</i>       | <i>Coefficient</i> | <i>Standard Error</i> |
|------------------------|--------------------|-----------------------|
| $\Delta \ln TL_t$      | <b>0.125**</b>     | <b>0.050</b>          |
| $\Delta \ln RDL_t$     | <b>0.068***</b>    | <b>0.023</b>          |
| $\Delta \ln IEL_t$     | <b>0.125**</b>     | <b>0.064</b>          |
| $\Delta \ln IEL_{t-1}$ | <b>-0.099*</b>     | <b>0.060</b>          |
| $\Delta \ln IEL_{t-2}$ | <b>-0.158***</b>   | <b>0.055</b>          |
| $\Delta C$             | <b>0.109</b>       | <b>0.108</b>          |
| $ecm(-1)$              | <b>-0.271***</b>   | <b>0.066</b>          |

Note 1: Dependent variable is  $\Delta \ln YL_t$

Note 2:  $ecm = \ln YL - 0.462 \ln TL - 0.251 \ln RDL - 0.208 \ln IEL - 0.402 C$

Note 3:  $R^2 = 0.65$  and  $\bar{R}^2 = 0.56$

Note 4: \*\*\*, \*\*, and \* denote significance at 1%, 5%, and 10%, respectively.

All of the short-run coefficients are significant at the 10% level and, except for the lagged values of  $\ln IEL$ , all of the coefficients are positive. This is in line with a priori expectations. The error correction term is negative and significant at the 1% level. The value of the correction factor is -0.27, meaning that 27% of any disequilibrium will be corrected annually.

In the short run, the largest proportion of change in labour productivity is explained by the stock of public infrastructure ( $\ln TL$ ): a 1% increase in the stock of public infrastructure per hour worked will increase labour productivity in the present period by 0.125%. The coefficients of  $\ln RDL$  and  $\ln IEL$  are 0.068% and 0.124%, respectively. The lagged coefficients of  $\ln IEL$ , meaning  $\ln IEL1$ , and  $\ln IEL2$  are negative, indicating that economic openness will negatively affect labour productivity in the two subsequent years.

The results of diagnostic tests also confirm that the specification of the model is correct. These diagnostic tests are: the Lagrange Multiplier (LM) test for serial correlation, the normality test, the ARCH test, RESET test and the unit root test of the residuals. The CUSUM and CUSUMQ tests were used to test for the stability of the coefficients of the model, and the model is found strictly stable as the statistics stay well within the critical bounds at 5% level of significance (Annex B).

Figure 2 depicts the actual and fitted values of  $\ln YL$  for the long-run model. The estimated model fits perfectly with the realized data, as can be seen below.

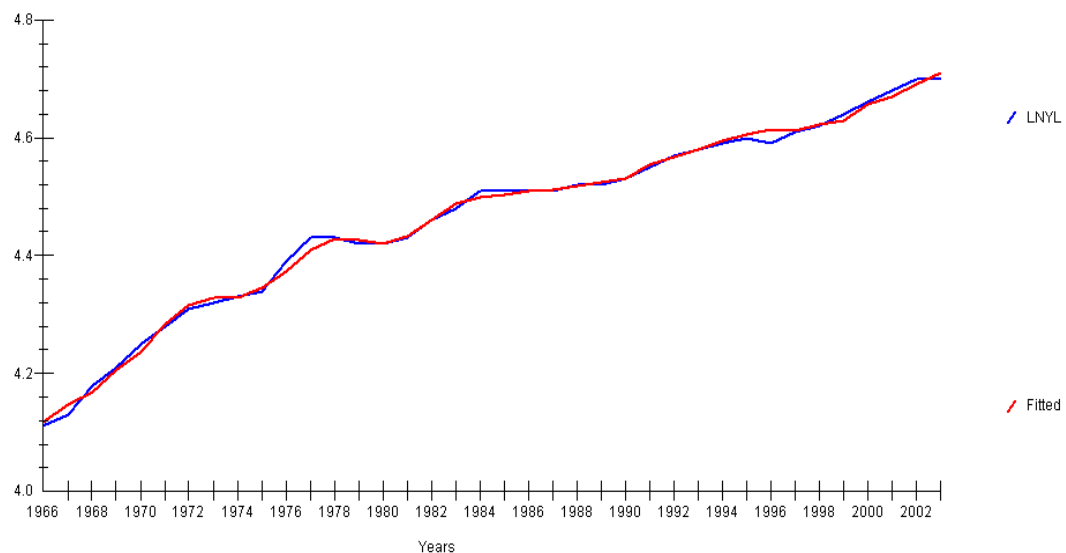


**Figure 2. Actual and fitted values for  $\ln YL$  in the long-run model**

Figure 3 shows the actual and fitted values of  $\ln YL$  for the error correction model. Again, the estimated model corresponds to the actual data.

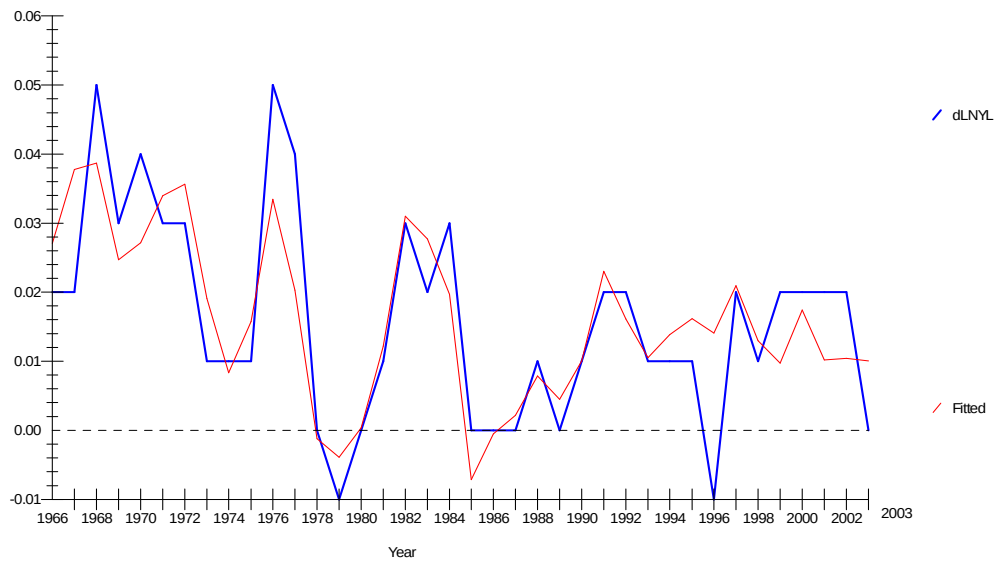


**Figure 3. Actual and fitted values for  $\ln YL$  in the error correction model**

The robustness of the results just stated is tested using the DOLS approach to cointegration. In performing DOLS estimation a maximum of 3 leads and lags of the first differences are used. The reported standard errors are based on the Newey and West (1987) adjustment, with a truncation lag of 3. The individual lead and lag parameters lack any significance and therefore economic interpretation, but the inclusion of the first-differenced variables is required to eliminate small-sample bias resulting from any correlation between the error term and the  $I(1)$  variables. The long-run coefficients suggest a broad similarity with those obtained using the ARDL approach (Table 3).

**Table 3.**

**Estimated long run coefficients using the DOLS approach**

| <i>Regressor</i> | <i>Coefficient</i> | <i>Standard Error</i> |
|------------------|--------------------|-----------------------|
| $\ln TL$         | <b>0.446**</b>     | <b>0.180</b>          |
| $\ln RDL$        | <b>0.094*</b>      | <b>0.053</b>          |
| $\ln IEL$        | <b>0.314**</b>     | <b>0.138</b>          |
| $C$              | <b>2.940**</b>     | <b>1.196</b>          |

Note 1: Dependent variable is  $\ln YL$

Note 2: \*\*\*, \*\*, and \* denote significance at 1%, 5%, and 10%, respectively.

## **1.5 Conclusion**

The paper, in a rigorous manner, defines public infrastructure and then uses the definition to tabulate the time series to represent the variable. Observations on the variable, along with those for labour productivity, economic openness, R&D expenditure, and total hours worked over a forty-three-year period are then used to estimate an econometric model and examine the sources of labour productivity growth in the Canadian economy. In the short run all three major explanatory variables – the stock of public infrastructure, economic openness and R&D expenditure – show positive effects on labour productivity. The coefficients of the first and second lagged values of trade openness, however, are indicative of a negative relationship between mid-term economic openness and labour productivity. A negative correction factor confirms an adjustment period of approximately four years for any deviation from the equilibrium status of labour productivity.

The strong relationship between labour productivity and the stock of public infrastructure suggests that the latter is the main source of long-run productivity growth. R&D costs and economic openness were also found to have a positive long-run relationship with labour productivity.

The results suggest that, growth-wise, Canada benefited from the boom in capital formation that occurred in late 1980s and throughout the 1990s. Trade exposure, although important in scope, did not likely play as important a role in the Canadian context as it did in the United States. Improvements in trade exposure, along with a shift in its composition toward more infrastructure-oriented services may help to spur growth in the long run. In terms of R&D intensity, Canada's performance has improved over time, particularly due to improvements in business sector participation however its overall position among other advanced economies, especially the United States, still falls behind. This is partly attributable to Canada's industrial orientation i.e. having an economy heavily dependent on natural resources, and service industries. The recent attention to changing the R&D accounting treatments in Canada, as well as an increasingly focused attention to other influencing factors, such as competition and patent policies, may help restore the country's international position in the long run. The next two chapters of this dissertation will tease out the subjects theoretically in an attempt to bring more clarity to this issue.

The findings of this chapter are therefore along the lines of neoclassical growth theory insofar as accumulation of infrastructure investment explains the majority of growth in labour productivity (see Solow 1956 & 1957 for more detail). The remainder of the productivity growth in the Canadian context is explained by technological progress resulting from R&D expenditures as well as economic openness, as suggested by the new growth theory (see Romer, 1986, 1987 & 1990, Lucas, 1988 and Grossman and Helpman, 1991 for more detail).

An interesting addition to the current work would be the incorporation of a provincial/territorial analysis of public infrastructure to the context, as this could provide significant information on the drivers of regional growth and prosperity. Data limitations have been the main preventing factor in leaving this task unaccomplished. It is also acknowledged that the effect of public infrastructure on productivity may be different in economic recession and boom periods. An attempt to control for these

differences and to distinguish between these impacts would help to substantiate the findings of this paper.

## Annex A

The ARDL method provides a simple framework to obtain a dynamic ECM. This could be illustrated in form of a simplest two-variable ARDL model as follows:

$$(A.1) \quad Y_t = \alpha + \sum_{i=1}^n \mu_i Y_{t-i} + \sum_{i=0}^m \eta_i X_{t-i} + u_t .$$

In order to trace a long-run relationship, consider the steady-state point of the model in which  $Y_t$  and  $X_t$  will be equal to their steady-state levels  $Y^*$  and  $X^*$ , respectively. Therefore, we have

$$(A.2) \quad Y^* = Y_t = Y_{t-1} = Y_{t-2} = \dots = Y_{t-n} ,$$

$$X^* = X_t = X_{t-1} = X_{t-2} = \dots = X_{t-m} .$$

Substituting (A.2) into (A.1), we obtain the following long-run equation:

$$(A.3) \quad Y^* = \frac{\alpha}{1 - \sum \mu_i} + \frac{\sum \eta_i}{1 - \sum \mu_i} X^* ,$$

or simply,

$$(A.4) \quad Y^* = \theta_0 + \theta_1 X^* .$$

and the equilibrium error term ( $e_t$ ) as

$$(A.5) \quad e_t = Y_t - Y^* = Y_t - \theta_0 - \theta_1 X_t .$$

Using OLS we can estimate  $\theta_0$  and  $\theta_1$ . We can also use a simple linear transformation of the model to obtain an ECM representation of the model. Reparametrizing (A.1), we have

$$(A.6) \quad \Delta Y_t = \alpha + \sum_{i=1}^{n-1} \mu_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \eta_i \Delta X_{t-i} + \vartheta_1 Y_{t-1} + \vartheta_2 X_{t-1} + \varepsilon_t.$$

In this model  $\vartheta_2 = \sum_{i=1}^m \eta_i$  (the numerator of  $\theta_1$ ) and  $\vartheta_1 = -(1 - \sum_{i=1}^n \mu_i)$ . So the ECM will have the following form:

$$(A.7) \quad \Delta Y_t = \alpha + \sum_{i=1}^{n-1} \mu_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \eta_i \Delta X_{t-i} + \vartheta_1 (Y_{t-1} - \frac{1}{\vartheta_1} - \frac{\vartheta_2}{\vartheta_1} X_{t-1}) + \varepsilon_t,$$

or

$$(A.8) \quad \Delta Y_t = \alpha + \sum_{i=1}^{n-1} \mu_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \eta_i \Delta X_{t-i} - \tau (Y_{t-1} - \hat{\theta}_0 - \hat{\theta}_1 X_{t-1}) + \varepsilon_t.$$

Here we have used the fact  $\tau = -\vartheta_1$ . In addition, we know that  $Y_{t-1} - \hat{\theta}_0 - \hat{\theta}_1 X_{t-1} = \hat{e}_{t-1}$ .

Therefore, the above ECM could be rewritten as

$$(A.9) \quad \Delta Y_t = \alpha + \sum_{i=1}^{n-1} \mu_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \eta_i \Delta X_{t-i} - \tau \hat{e}_{t-1} + \varepsilon_t.$$

In (A.9),  $\tau$  is the adjustment coefficient or error correction coefficient.

## Annex B: Test Results

Table B.1. Autoregressive Distributed Lag Estimates

| <i>Regressor</i> | <i>Coefficient</i> | <i>t-Ratio</i> |
|------------------|--------------------|----------------|
| <i>LNYL(-1)</i>  | 0.729***           | 11.019         |
| <i>LNTL</i>      | 0.125**            | 2.523          |
| <i>LNRDL</i>     | 0.068***           | 2.932          |
| <i>LNIEL</i>     | 0.125*             | 1.969          |
| <i>LNIEL(-1)</i> | -0.168*            | -1.712         |
| <i>LNIEL(-2)</i> | -0.059             | -0.631         |
| <i>LNIEL(-3)</i> | 0.158***           | 2.879          |
| <i>C</i>         | 0.109              | 1.008          |

Note 1: Dependent variable is  $\ln YL$

Note 2: \*\*\*, \*\*, and \* denote significance at 1%, 5%, and 10%, respectively.

Note 3:  $\bar{R}^2 = 0.99$

Note 4: F-statistic is equal to 1377.9, and Durbin's h-statistic is equal to 1.349

Table B.2. Diagnostic Tests (ARDL Approach)

| <i>Test Statistics</i>       | <i>LM Version</i>       | <i>F Version</i>        |
|------------------------------|-------------------------|-------------------------|
| <i>A: Serial Correlation</i> | CHSQ(1) = 1.378 [0.240] | F(1,29) = 1.091 [0.305] |
| <i>B: Functional Form</i>    | CHSQ(1) = 2.434 [0.119] | F(1,29) = 1.985[0.170]  |
| <i>C: Normality</i>          | CHSQ(2) = 0.627[0.731]  | Not applicable          |
| <i>D: Heteroscedasticity</i> | CHSQ(1) = 0.409 [0.522] | F(1, 36) = 0.393[0.535] |

A: Lagrange multiplier test of residual serial correlation

B: Ramsey's RESET test using the square of the fitted values

C: Based on a test of skewness and kurtosis of residuals

D: Based on the regression of squared residuals on squared fitted values

Note: The values reported inside the brackets are p-values

**Table B.3. Error Correction Representation of the Selected ARDL Model**

| <i>Regressor</i> | <i>Coefficient</i> | <i>t-Ratio</i> |
|------------------|--------------------|----------------|
| <i>dLNTL</i>     | <b>0.125**</b>     | <b>2.523</b>   |
| <i>dLNRDL</i>    | <b>0.068***</b>    | <b>2.932</b>   |
| <i>dLNIEL</i>    | <b>0.125**</b>     | <b>1.969</b>   |
| <i>dLNIEL1</i>   | <b>-0.099*</b>     | <b>-0.667</b>  |
| <i>dLNIEL2</i>   | <b>-0.158***</b>   | <b>-2.879</b>  |
| <i>dC</i>        | <b>0.109</b>       | <b>1.008</b>   |
| <i>ecm(-1)</i>   | <b>-0.271***</b>   | <b>-0.098</b>  |

Note 1: Dependent variable is  $\ln YL$

Note 2: \*\*\*, \*\*, and \* denote significance at 1%, 5%, and 10%, respectively.

Note 3: d is the difference operator.

Note 4:  $dLNIEL1 = LNIEL(-1) - LNIEL(-2)$  &  $dLNIEL2 = LNIEL(-2) - LNIEL(-3)$

$ecm = LNYL - 0.462*LNTL - 0.251*LNRDL - 0.208*LNIEL - 0.402*C$

Note 5:  $\bar{R}^2 = 0.56$

Note 6:  $F = 9.102$

**Table B.4. Estimated Long Run Coefficients (ARDL Approach)**

| <i>Regressor</i> | <i>Coefficient</i> | <i>t-Ratio</i> |
|------------------|--------------------|----------------|
| <i>LNTL</i>      | <b>0.462***</b>    | <b>3.990</b>   |
| <i>LNRDL</i>     | <b>0.251***</b>    | <b>4.994</b>   |
| <i>LNIEL</i>     | <b>0.208**</b>     | <b>2.379</b>   |
| <i>C</i>         | <b>0.402</b>       | <b>0.958</b>   |

Note: Dependent variable is  $\ln YL$

**Table B.5. Test of Serial Correlation of Residuals (OLS case)**

| <i>Regressor</i>    | <i>Coefficient</i> | <i>t-Ratio</i>        |
|---------------------|--------------------|-----------------------|
| <b>OLS RES(- 1)</b> | <b>0.071</b>       | <b>0.317[0.753]</b>   |
| <b>OLS RES(- 2)</b> | <b>-0.016</b>      | <b>-0.066[0.948]</b>  |
| <b>OLS RES(- 3)</b> | <b>-0.376*</b>     | <b>-1.826 [0.077]</b> |
| <b>OLS RES(- 4)</b> | <b>-0.231</b>      | <b>-1.007 [0.321]</b> |

Note 2: LM statistic: CHSQ = 7.0026[0.136]

Note 2: F-Statistic =1.4684[0.240]

**Table B.6. Unit Root Tests for Residuals Based on ARDL Regression of LNYL on:**

| <b>LNYL(-1)</b>       | <b>LNTL</b>   | <b>LNRD</b>    | <b>LNIEL</b>  | <b>LNIEL(-1)</b> | <b>LNIEL(-2)</b> | <b>LNIEL(-3)</b> | <b>C</b> |
|-----------------------|---------------|----------------|---------------|------------------|------------------|------------------|----------|
| <i>Test Statistic</i> | <i>LL</i>     | <i>AIC</i>     | <i>SBC</i>    | <i>HQ</i>        | <i>C</i>         |                  |          |
| <b>DF</b>             | <b>-4.103</b> | <b>98.109</b>  | <b>97.109</b> | <b>96.425</b>    | <b>96.898</b>    |                  |          |
| <b>ADF(1)</b>         | <b>-3.377</b> | <b>98.151</b>  | <b>96.151</b> | <b>94.783</b>    | <b>95.723</b>    |                  |          |
| <b>ADF(2)</b>         | <b>-3.554</b> | <b>99.008</b>  | <b>96.008</b> | <b>93.957</b>    | <b>95.366</b>    |                  |          |
| <b>ADF(3)</b>         | <b>-3.462</b> | <b>99.513</b>  | <b>95.513</b> | <b>92.778</b>    | <b>94.656</b>    |                  |          |
| <b>ADF(4)</b>         | <b>-3.757</b> | <b>100.821</b> | <b>95.821</b> | <b>92.403</b>    | <b>94.751</b>    |                  |          |
| <b>ADF(5)</b>         | <b>-3.352</b> | <b>101.099</b> | <b>95.099</b> | <b>90.997</b>    | <b>93.814</b>    |                  |          |
| <b>ADF(6)</b>         | <b>-1.895</b> | <b>103.045</b> | <b>96.045</b> | <b>91.259</b>    | <b>94.546</b>    |                  |          |
| <b>ADF(7)</b>         | <b>-1.303</b> | <b>104.110</b> | <b>96.110</b> | <b>90.641</b>    | <b>94.397</b>    |                  |          |
| <b>ADF(8)</b>         | <b>-1.535</b> | <b>104.839</b> | <b>95.839</b> | <b>89.686</b>    | <b>93.912</b>    |                  |          |

LL = Maximized log-likelihood      AIC = Akaike Information Criterion

SBC = Schwarz Bayesian Criterion      HQC = Hannan-Quinn Criterion

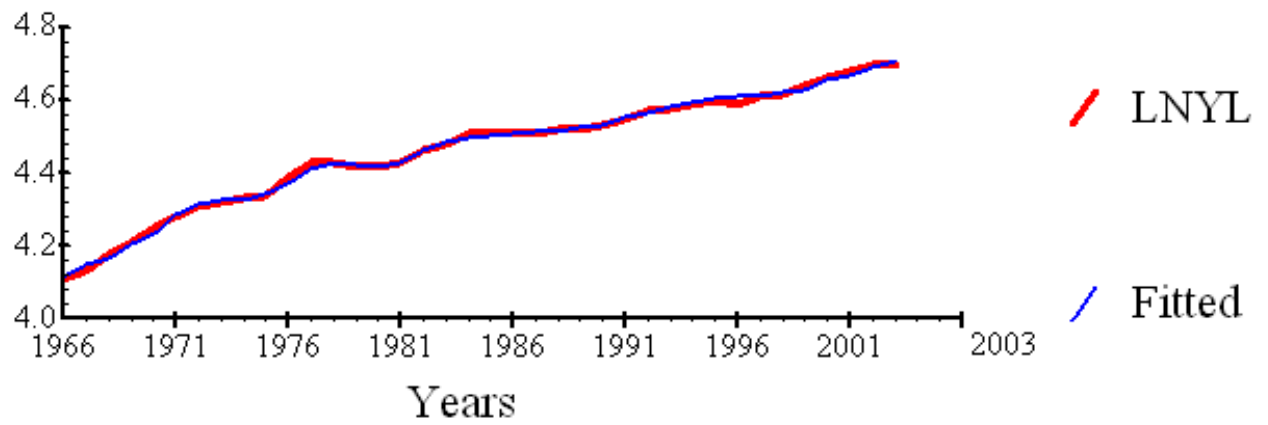


Figure B.1. Actual and Fitted Values for  $\ln YL$  (Sample from 1966 to 2003)

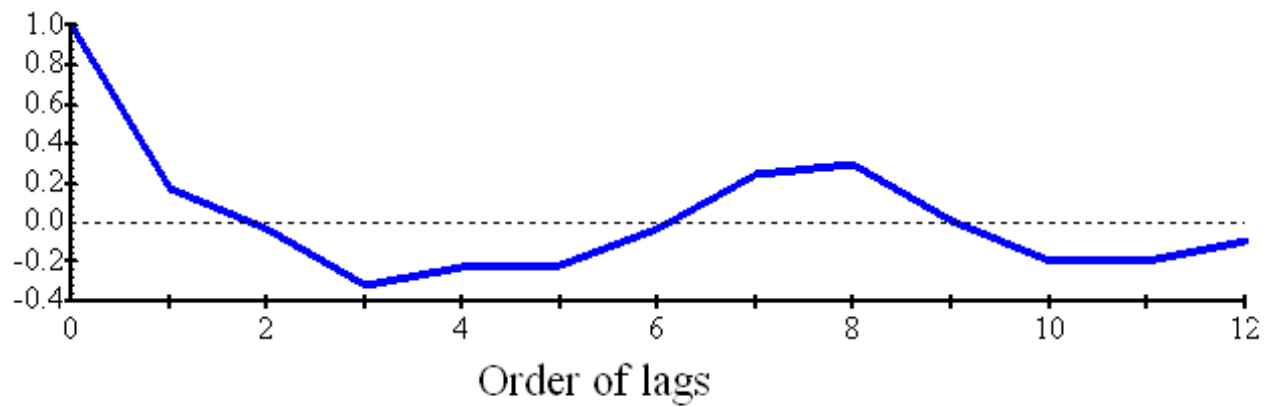


Figure B.2. Autocorrelation Function of Residuals (Sample from 1966 to 2003)

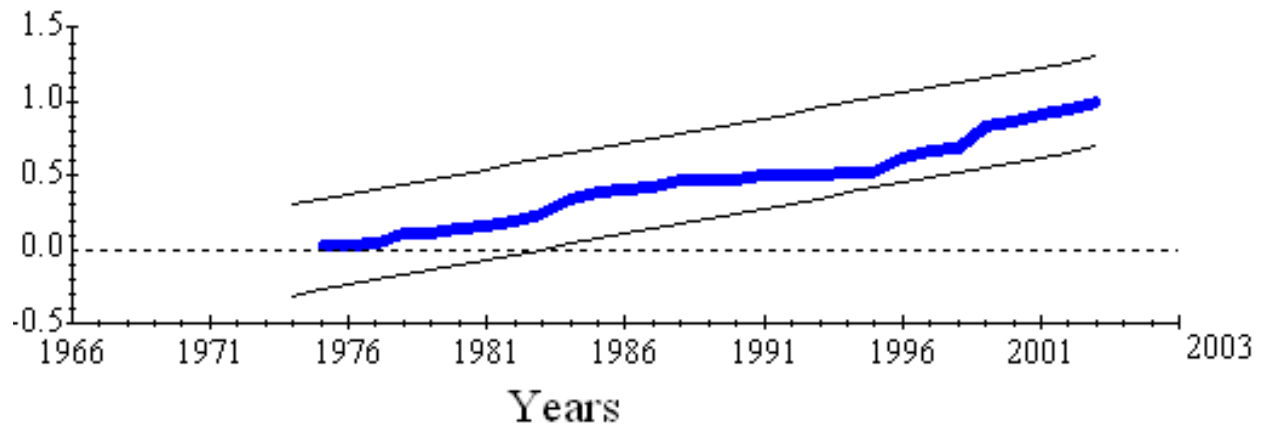


Figure B.3. Cumulative Sum of Squares of Recursive Residuals (Note: The straight lines represent critical bounds at 5% significance level)

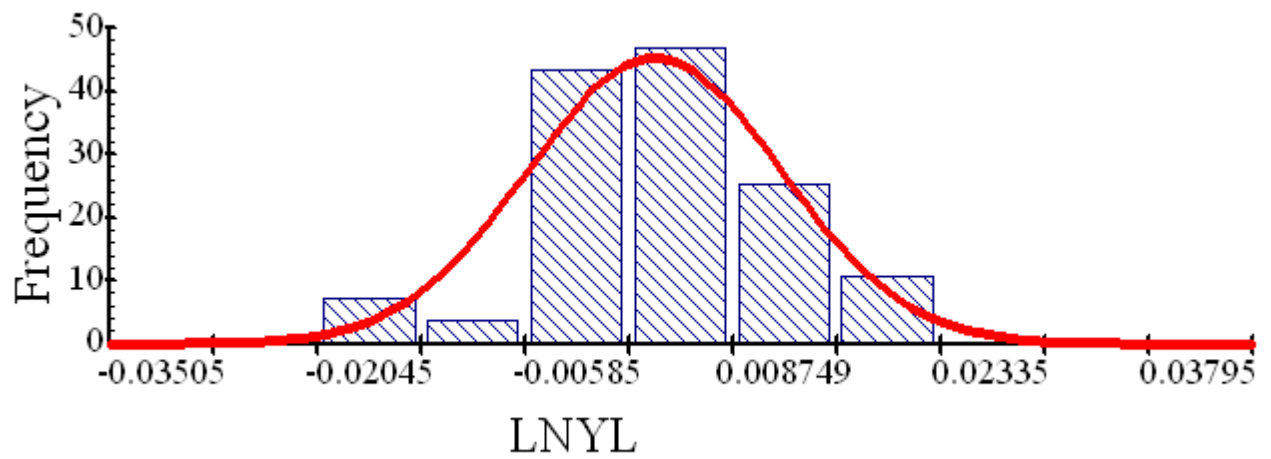


Figure B.4. Histogram of Residuals and the Normal Density (Sample from 1996 to 2003)

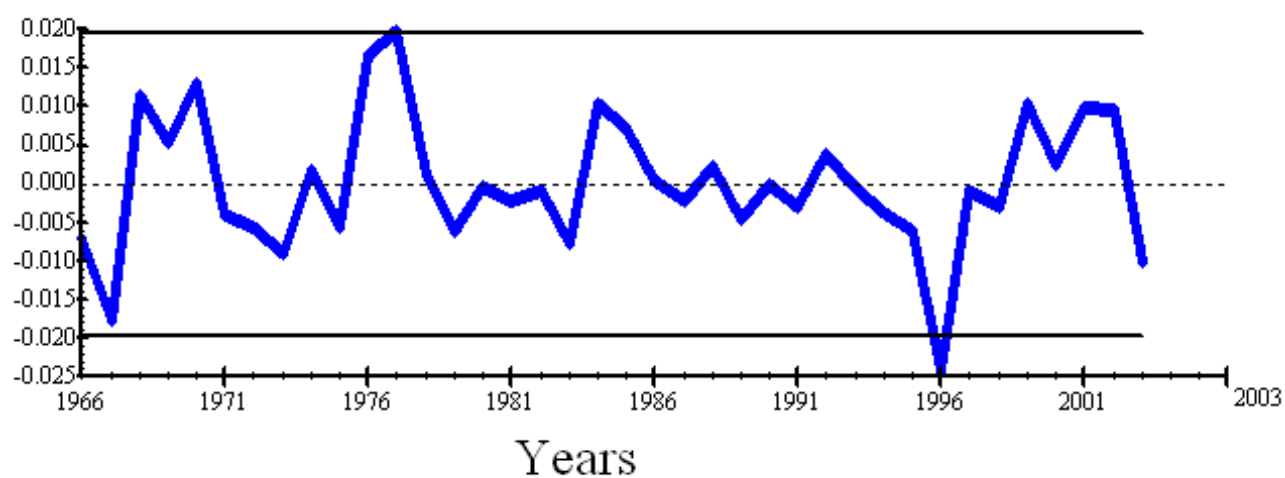


Figure B.5. Residuals and Two Standard Error Bands (Sample from 1996 to 2003)

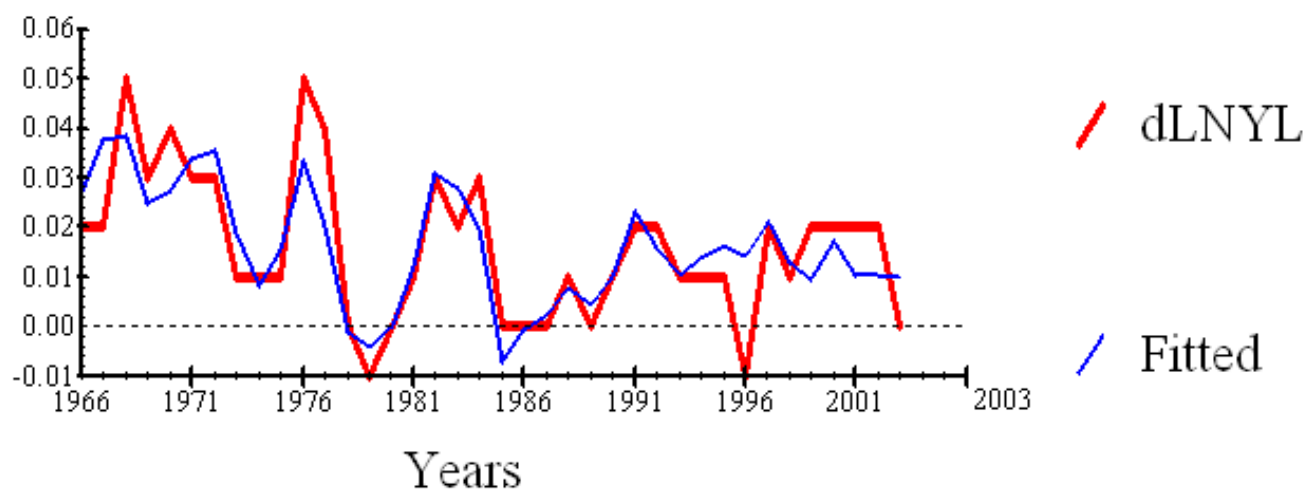


Figure B.6. Actual and Fitted Values of  $\ln YL$  in the Error Correction Model (Sample from 1996 to 2003)

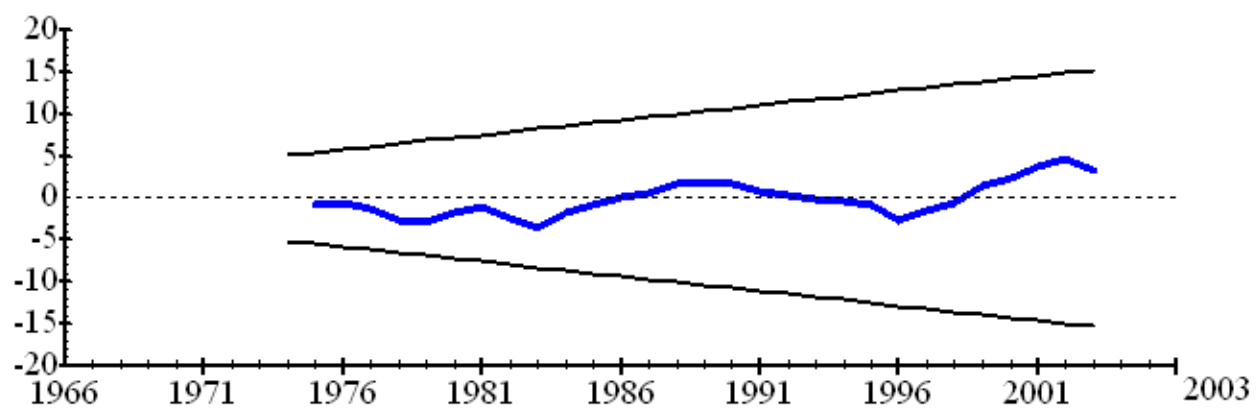


**Figure B.7. Cumulative Sum of Recursive Residuals (Note: The straight lines represent critical bounds at 5% significance level)**

## References

- Abdi, T. (2008)**, "Machinery and Equipment Investment and Growth: Evidence from the Canadian Manufacturing Sector," *Applied Economics*, 40(4-6), 465-78
- Aschauer, D. A. (1989)**, "Is Public Expenditure Productive," *Journal of Monetary Economics*, 23, 177-200.
- Bin Dong, L. (2005)**, "The Private Benefits of Public Capital: Evidence for Canada, 1987 to 2002," *Economic Analysis Research Paper Series*, Statistics Canada.
- Blomstrom, M.; R. Lipsey, and M. Zejan (1996)**, "Is Fixed Investment the Key to Economic Growth," *Quarterly Journal of Economics*, 111, 269-76.
- Brox, J. and C. Fader (2005)**, "Infrastructure Investment and Canadian Manufacturing," *Applied Economics*, 37, 1247-56.
- Clarida, R. (1993)**, "International Capital Mobility, Public Investment and Economic Growth," *NBER Working Papers 4506*, National Bureau of Economic Research Inc.
- De Long, J and L. Summers (1991)**, "Equipment Investment and Economic Growth," *The Quarterly Journal of Economics*, MIT Press, 106(2), 445-502.
- Domar, E.D. (1946)** "Capital Expansion, Rate of Growth, and Employment," *Econometrica*, 14, 137-147.
- Dowrick, S. and J. Golly (2004)**, "Trade Openness and Growth: Who Benefits?" *Oxford Review of Economic Policy*, 20 (1), 38-56.
- Frankel, J. F. and D. Romer (1999)**, "Does Trade Cause Growth?" *The American Economic Review*, 89(3), 379-399.
- Gera, S.; W. Gu, and F. Lee (1999)**, "Information Technology and Labour Productivity Growth: An Empirical Analysis for Canada and the United States," *Canadian Journal of Economics*, 32(2), 384-407.
- Grossman, G. M. and E. Helpman (1991)**, "Innovation and Growth in the Global Economy," *MIT Press*, Cambridge.
- Gramlich, E. M. (1994)**, "Infrastructure Investment: A Review Essay," *Journal of Economic Literature*, XXXII, 1176-1196.
- Harchaoui, T. and F. Tarkhani (2003)**, "Public capital and its contribution to the productivity performance of the Canadian business sector," *Catalogue no. 11F0027MIE - No.017*, Statistics Canada.
- Harrod, R. F. (1939)**, "An Essay in Dynamic Theory," *Economic Journal*, 49, 14-33.

- Khazabi, M. (2008)**, "On R&D investment," *ICU Journal of Managerial Economics*, 6(2), 54-68.
- Koehler, A and E. Murphree (1988)**, "A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37(2), 187-195
- Lileeva, A. (2008)**, "Trade Liberalization and Productivity Dynamics: Evidence from Canada," *Canadian Journal of Economics*, 41(2), 360-90.
- Lucas, R. E. (1988)**, "On the Mechanics of Development Planning," *Journal of Monetary Economics*, 22, 3-42.
- Levine, R. and D. Renelt (1992)**, "A Sensitivity Analysis of Cross-Country Growth Regressions," in *Economic growth in the long run: A history of empirical evidence*, 2, 270-91.
- Lynde, C. and J. Richmond (May 1993)**, "Public Capital and Total Factor Productivity," *International Economic Review*, 401-14
- Madden, G. and S. J. Savage (1998)**, "Sources of Australian Labour Productivity Change 1950-1994," *The Economics Record*, 74 (227), 362-72.
- Macklem, T. (2003)**, "Future Productivity Growth in Canada: Comparing to the United States," *International Productivity Monitor*, 7, 50-57.
- Mittnik, S. and T. Neumann (2000)**, "Dynamic Effects of Public Investment: Vector Autoregressive Evidence from Six Industrialized Countries." *Empirical Economics*, 26, 429-46.
- Newey, W. & K. West (1987)**, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Econometric Society, 55(3), 703-08
- OECD (2001)**, *Measuring Productivity: Measurement of Aggregate and Industry-level productivity growth*, *OECD Manual*, OECD, Paris.
- Otto, R.M. and G. Voss (2003)**, "Public investment and economic growth," *Applied Economics*, 35, 527-540.
- **(1998)**, "Is public capital provision sufficient?" *Journal of Monetary Economics*, 42, 47-66.
- **(1996)**, "Public Capital and Private Production in Australia," *Southern Economic Journal*, 62, 723-38

- Pesaran, H.; Y. Shin and R. Smith (2001)**, “Bounds testing approaches to the analysis of level relationships,” *Journal of Applied Econometrics*, 16, 289-326.
- Pesaran, H. and B. Pesaran (1997)**, *Working with Microfit 4.0: Interactive Econometric Analysis*, London: Oxford University Press.
- Pesaran, H. and Y. Shin (1995)**, “An Autoregressive Distributed Lag Modeling Approach to Cointegration Analysis”, in *Storm, S., Holly, A. and Diamond, P. (eds.) Centennial Volume of Rangar Frisch, Econometric Society Monograph*, Cambridge: Cambridge University Press.
- Ratner, J. B. (1983)**, “Government capital and the production function for US private output,” *Economic Letters*, 213-17.
- Rao, Someshwar; A. Ahmad; W. Horsman, and P. Kaptein-Russell (2001)**, “The Importance of Innovation for Productivity,” *International Productivity Monitor*, 2, 11-18.
- Rao, S.; A. Sharpe, and J. Smith (2005)**, “An Analysis of the Labour Productivity Growth Slowdown in Canada since 2000,” *International Productivity Monitor*, 10, 3-23.
- Rao, S.; J. Tang, and W. Wang (2003)**, “Canada's Recent Productivity Record and Capital Accumulation,” *International Productivity Monitor*, 7, 24-38.
- Romer, P. (1986)**, “Increasing returns and long – run growth,” *The Journal of Political Economy*, 94(5), 1002-1037.
- **(1987)**, "Growth Based on Increasing Returns Due to Specialization," *American Economic Review*, 77(2), 56-62
- **(1990)**, "Are Nonconvexities Important for Understanding Growth?," *American Economic Review*, 80(2), 97-103
- Sachs J. D. and A. M. Warner (1997)**, "Fundamental Sources of Long Run Growth," *American Economic Review*, 87 (2), 84-188.
- Satya P.; S. Balbir, and B. Bagala (2004)**, “Public Infrastructure and the Productive Performance of Canadian Manufacturing Industries,” *Southern Economic Journal*, 70 (4), 998-1011.
- Shell, K. (1967)** "Optimal Programs of Capital Accumulation for an Economy in which there is Exogenous Technical Change", in *Essays on the Theory of Optimal Economic Growth*. Cambridge, M.I.T. Press.
- Solow, R. M. (1956)**, “A Contribution to the Theory of Economic Growth,” *Quarterly Journal of Economics*, 70, 65-94.

- Torrisi, G (2009)**, "Public infrastructure: definition, classification and measurement issues," *MPRA Paper 12990*, University Library of Munich, Germany.
- Trefler, D. (2004)**, "The Long and Short of the Canada-U.S. Free Trade Agreement," *STICERD - Economics of Industry Papers*, LSE.
- Wylie, P. (1996)**, "Infrastructure and Canadian Economic Growth 1946-1991," *Canadian Journal of Economics*, 29, 350-355.

## **2 R&D Spillovers, Innovation, and Entry**

### **2.1 Introduction**

In the economic literature, innovation – also called R&D – is often classified as product innovation or process innovation. A firm carries out a program of product innovation to find a new product that it hopes will generate new demand and lead to large profits. Process innovation, on the other hand, aims at finding a new process to reduce the production cost of a product. A lower production cost, which is the desired outcome of the R&D program, gives the firm a cost advantage over its rivals. Whether a program of innovation will be carried out or not depends on the cost of R&D and the market structure in which the firm finds itself. Knowledge and benefits obtained by a firm from its R&D activities typically leak out to other firms, to consumers and, eventually, to other countries. These leakages – called R&D spillovers – mean that a firm cannot appropriate all the fruit of its R&D activities, especially when spillovers flow to its competitors in the same industry. However, from society's point of view, spillovers represent positive externalities in the sense that they reduce the production costs of other firms, with the ensuing consequence of lower prices for consumers. In a review of the literature on R&D aimed at providing guidelines for recent efforts to include R&D in the national income accounts, Sveikauskas (2007) indicated that perhaps the private rate of return to R&D is 25%, while it is 65% for social returns.

In light of the positive externalities generated by R&D activities, the authorities charged with competition policy in Europe and Japan have adopted a rather permissive anti-trust attitude toward R&D cooperation for quite some time. The research – both theoretical and empirical – received the needed impetus in 1984 when

the US passed the National Cooperation Act in 1984, allowing firms to cooperate in R&D, but not in product markets. Over the last two decades, the economics of R&D spillovers has been one of the most active fields of research in industrial economics.

The theoretical literature on competition and cooperation in R&D with technological spillovers can be said to begin with the pioneering work of d'Aspremont and Jacquemin (1988) (AJ hereafter), who formulated a two-stage duopoly game of R&D spillovers in which the two firms behave in a non-cooperative manner in the second (production) stage, but can either cooperate or behave in a non-cooperative manner in the first (R&D) stage. Now when the two firms behave cooperatively in the R&D stage, it is reasonable to expect that the cooperation will lead to a lower level of total R&D expenditures made by the two firms because of less wasteful duplications and a lower level of total output resulted from the monopoly power. AJ demonstrated that these expectations are far from being fulfilled in a simple two-stage linear-quadratic game – linear demand curve, linear total cost, and quadratic R&D costs. Three different scenarios of competition are considered by AJ. Under the first scenario, the two firms act non-cooperatively in both stages of the game. Under the second scenario, the two firms cooperate in the first stage, but behave non-cooperatively in the second stage. Under the third scenario, the two firms behave jointly like a single integrated firm in both stages of the game. AJ also considered the problem faced by the central planner. The linear-quadratic structure of the model makes it possible to obtain a closed-form solution for each of the problems and allows for a comparison of the solutions – R&D expenditures and welfare – of the four models. In the analysis carried out by AJ, the degree of R&D spillovers plays a critical role. AJ showed that when the degree of R&D spillovers is high, the level of R&D expenditures and total output are higher under the second scenario than under the first scenario. Otherwise, the opposite results hold. For a high degree of R&D spillovers, the R&D expenditures under the social optimum are highest to be followed – in descending order – successively by the R&D expenditures under the scenario that the two firms act like a single integrated firm, the scenario that the two firms cooperate in R&D, but act non-

cooperatively in the production stage, and the scenario that the two firms behave non-cooperatively in both stages of the game.

The results on the comparative performance of non-cooperative and cooperative R&D derived by AJ have received – in the duopoly context – a thorough generalization by Amir et al. (2003). In particular, when the two firms cooperate in R&D, these researchers allowed the firms to determine jointly their R&D expenditures and the degree of R&D spillovers. The R&D degree of spillovers is thus endogenous, and can be chosen to maximize joint profits net of R&D costs.

The AJ model was extended by Suzumura (1992) to the case of many firms, general demand, and general cost conditions. The more general model of Suzumura precludes the possibility of computing the equilibria of the two-stage games for various specifications of R&D, and it is no longer possible to compare these equilibria directly. This researcher resolved this difficulty by trying to answer the question of starting from an equilibrium – under non-cooperative R&D or cooperative R&D – can social welfare be raised by marginally increasing R&D expenditures? Two measures of social welfare are used in these exercises: the first-best social optimum and the second-best social optimum. According to the first-best measure of social welfare, the sum of consumer and producer surplus is maximized, and the marginal cost pricing rule, which underlines the first-best solution, can be enforced by the authorities. According to the second-best measure of social welfare, the firms are allowed to compete according to the Cournot model of competition. Suzumura demonstrated that when the degree of spillovers is high, starting from the equilibrium level of R&D expenditures under the scenario that the firms act non-cooperatively in both stages of the game, first-best social welfare can be raised by marginally increasing R&D expenditures. The opposite result holds if there are no spillovers. As for the equilibrium under cooperative R&D, first-best social welfare can be raised by marginally increasing R&D expenditures, whether the degree of spillovers is high or low. If one uses the second-best measure of social welfare, then starting from the equilibrium level of R&D expenditures under the scenario that the firms act non-

cooperatively in both stages of the game, social welfare can be raised by marginally increasing R&D expenditures if the degree of spillovers is high. The opposite result holds if there are no R&D spillovers and if the number of firms is sufficiently large. As for the case of cooperative R&D, social welfare can be raised by marginally increasing R&D expenditures, whether the degree of spillovers is high or low.

The welfare results of Suzumura are obtained under two extreme assumptions – high and low degrees of spillovers. Yi (1996) completed the analysis of Suzumura by considering the intermediate case of neither high nor low degrees of spillovers. More specifically, Yi established the following results. First, cooperative R&D lowers both R&D expenditures and social welfare for intermediate degrees of spillovers. Second, cooperative R&D lowers R&D expenditures, but has an ambiguous effect on social welfare for low degrees of spillovers. Third, as the elasticity of the slope of the inverse market demand curve rises, cooperative R&D raises social welfare for a larger set of degrees of spillovers, and in the limit, is socially beneficial for all degrees of spillovers.

The AJ's model has also been extended by other researchers, such as Kamien et al. (1992), Kamien and Zang (1993), Poyago-Theotoky (1996), and Atallah (2000) to study the issue of R&D cartelization and research joint ventures. In the models formulated by these authors a subset of the firms in the industry might get together and form a Research Joint Venture. A survey of the main results of the literature on spillovers and innovative activities is provided by De Bondt (1996). The predictions of the AJ model, especially the important question of whether spillovers increase firms' incentives to cooperate in R&D, has been addressed by a number of empirical studies with mixed results; see, for example, Cassiman and Veugelers (2002) and Sustens (2004).

In the literature on R&D spillovers and process innovation, efforts are mostly focused on the comparative R&D expenditures and the relative social welfare between non-cooperative and cooperative R&D. The question of how innovation is affected when

there is more competition, i.e., when the number of firms rises, is ignored by most researchers, except for De Bondt et al. (1992), who discussed how the number of firms affects innovation. However, the question of how R&D spillovers affect entry was not addressed by these researchers. In this paper, we attempt to fill part of this lacuna by endogenizing the number of firms. More specifically, our model addresses the following questions. First, how does the degree of spillovers affect the equilibrium number of firms? Second, how does the degree of spillovers affect the equilibrium market structure? Third, when is the equilibrium symmetric and when is it not symmetric, and in the case of asymmetric equilibria, how many firms choose to incur a positive amount of own R&D cost and how many firms choose to free ride on the R&D activities of others?<sup>5</sup>

The model we formulate to analyze the influence of R&D spillovers on entry is a two-stage game played by a number of firms producing a homogeneous good. In the first stage of the game, the firms carry out R&D activities to lower their production cost. It is assumed that before innovation, all the firms have the same marginal cost. In the second stage, the firms compete in the product market according to the Cournot model of competition. All the firms act non-cooperatively in both stages of the game. In the R&D stage, each firm runs its own research lab, and takes into account the natural spillovers that flow among firms in a strategic manner. In modeling the horizontal spillovers among firms, we follow the pioneering work of Ruff (1969), who analyzed a stylized growth model in which firms compete according to the Cournot model of competition, and in which firms undertake R&D activities by employing research workers. In Ruff's model, a firm recognizes a potential transmission of knowledge from other firms, and the transmission of knowledge is modeled by assuming that the *effective input* in R&D of a firm consists of its own input plus part of the inputs of all the other firms. Ruff's analytical treatment of R&D spillovers has been adopted by later researchers, such as Spence (1984), Kamien et al.

---

<sup>5</sup>See Dasgupta and Stiglitz (1980), Atallah (2007), Vives (2008), and Erkal and Piccinin (2010) for more information.

(1992), and Amir et al. (2003). On the other hand, in the AJ-type models, the R&D process is represented by a cost function, which gives the R&D cost (the dependent variable) as a quadratic function of the desired level of cost reduction (the independent variable). In these models, a firm chooses its own level of cost reduction, and the spillovers take the form of R&D output spillovers in the sense that the *effective cost reduction* of a firm is the firm's own chosen cost reduction plus a fraction of the cost reductions chosen by all the other firms. At first sight, one might think that the two ways of modeling R&D spillovers are equivalent, and should yield the same results. However, this presumption is not correct because for some questions the results obtained by the two approaches diverge. For example, Amir et al. (2003) found that a firm's *effective R&D expenditure* is a strictly decreasing function of the degree of spillovers while De Bondt et al. (1992) found an inverted-U relationship between the degree of spillovers and the *effective cost reduction* of each firm.

Our findings can be described as follows. If the degree of spillovers is low, only a finite number of firms enter the market, and after entry – all the firms expend the same amount of their own resources on R&D. We refer to this type of equilibrium as a symmetric equilibrium with innovation. The intuition behind this result is not hard to understand. When the degree of spillovers is low, a firm cannot rely on the R&D externalities generated by the other firms to lower its own marginal cost. If the degree of spillovers is high, an infinite number of firms enter the market, and after entry none of the firms chooses to expend any of its own resource on R&D. The equilibrium is a perfectly competitive equilibrium without innovation. For intermediate values of the degree of spillovers, the equilibrium is an asymmetric equilibrium under which some firms choose to expend their own resources on R&D, while others choose not to do so. Furthermore, all the firms that choose to expend their own resources on R&D choose the same amount of own expenditure on R&D. In the literature on R&D spillovers and innovation, the number of firms in the market is taken as exogenous, and the equilibrium is presumed to be symmetric. In our model, the number of firms that enter the market is endogenous, and varies according

to the degree of spillovers. The endogenization of the number of firms yields different types of equilibrium market structures – symmetric equilibrium with innovation, asymmetric equilibrium with innovation, and perfect competition without innovation – and these are novel results in the field of R&D spillovers and innovation.

Our welfare analysis suggests that social welfare rises with the degree of spillovers when it is low, reaches a maximum when the degree of spillovers enters its intermediate range, and then declines to the level associated with the competitive equilibrium without innovation. The policy implication of this result is that the intellectual property right should not be fully protected and the enforcement of patent laws should not be too lax. The optimal degree of protection should reflect the right trade-off between allocative and dynamic efficiency.

The paper is organized as follows. In Section 2, the general features of the model are presented. In Section 3, the post-innovation equilibrium is discussed. The equilibrium in the innovation stage is analyzed in Section 4. In Section 5, entry is discussed. Section 6 presents the main properties of the equilibrium. In Section 7, a numerical simulation of the model is presented. The simulation illustrates the main properties of the model. Some concluding remarks are given in Section 8. The annex contains some technical arguments used to support the theoretical arguments in Section 6.

## **2.2 The general features of the model**

Consider the market for a homogenous commodity in which there are  $n$  symmetric firms. These firms play a two-stage game, which can be described as follows. In the first stage, each firm carries out an R&D program to reduce its production cost. In the second stage, the firms – with lower marginal costs brought about by the process innovation programs carried out in the first stage – compete in the product market. Let

$$(1) \quad p = a - bQ$$

be the inverse market demand curve for the product, where  $Q$  is the industry output, and  $p$  is the market price. Also,  $a$  and  $b$  are two positive parameters.

We assume that in both stages of the game the firms act non-cooperatively and that in the second stage they compete according to the Cournot model of competition. In the first stage, and before the process innovation, the firms are symmetric in the sense that all the firms start with the same initial marginal cost, say  $c^0, 0 < c^0 < a$ . To model the R&D process, we follow Ruff (1969), and suppose that research workers constitute the only input in a program of process innovation. Furthermore, assuming that the wage received by a worker is the numéraire, we can identify the number of research workers with the R&D expenditure. We shall let  $f[X]$  denote the reduction in marginal cost yielded by a program of process innovation when  $X$  is the firms *effective R&D expenditure*, with the effective R&D expenditure being the sum of the firm's own R&D expenditure plus the spillovers from the other innovation programs. The R&D production function is assumed to be continuously differentiable, strictly increasing, and strictly concave. Furthermore,  $f[0] = 0, 0 < f'[0] < \infty$ , and  $\sup_X f[X] \leq c^0$ . The condition  $\sup_X f[X] \leq c^0$  ensures that the cost reduction is strictly less than  $c^0$ , regardless of the level of effective R&D input. Also, we shall assume that  $f'[0]$  is not too low to discourage a firm, when it is the only firm in the market, from investing in R&D to reduce its marginal cost. Note that the Inada condition  $\lim_{X \rightarrow \infty} f'[X] = 0$  follows from the concavity and the boundedness of the R&D production function. An R&D technology that satisfies these assumptions is

$$(2) \quad f[X] = c^0(1 - e^{-\gamma X}),$$

where  $\gamma > 0$  is a parameter that characterizes the productivity of the R&D technology.<sup>6</sup>

The cost reduction obtained by firm  $i$  is assumed to be given by  $f[x_i + \beta \sum_{j \neq i} x_j]$ , where  $0 \leq \beta \leq 1$ , is a parameter that represents the degree of spillovers from the R&D activities of all the other firms, and  $x_j, j = 1, \dots, n$ , is firm  $j$ 's own R&D expenditure. The expression  $\beta \sum_{j \neq i} x_j$  represents the spillovers to firm  $i$  from the R&D activities of all the other firms. The sum  $X_i = x_i + \beta \sum_{j \neq i} x_j$  thus represents the effective R&D expenditure of firm  $i$  in its own program of process innovation. When  $\beta = 0$ , there are no spillovers, and when  $\beta = 1$ , there are full spillovers. The intermediate case  $0 < \beta < 1$  corresponds to the situation of partial spillovers.

## 2.3 The post-innovation equilibrium

When each firm carries out its own process innovation program, the reduction in the marginal cost of a firm, say firm  $i$ , is  $f[x_i + \beta \sum_{j \neq i} x_j]$ , and its post-innovation marginal cost is given by

---

<sup>6</sup> In the AJ model, the R&D cost needed to achieve a level of marginal cost reduction is assumed to be a quadratic function of the R&D output. The production function that corresponds to such a cost function has the functional form  $f[X] = 2\sqrt{\frac{X}{\gamma}}$ , where  $\gamma > 0$  is a parameter that characterizes the productivity of the R&D process, with a lower value of  $\gamma$  representing a more productive technology. Although it makes the computation of an equilibrium simple, the R&D technology of the AJ model has some undesirable features. First, when the effective R&D expenditure is large, the reduction in marginal cost will exceed the initial marginal cost, rendering the post-innovation marginal cost negative, and this is absurd. Second, given that the initial marginal cost was the outcome of past R&D activities and given the presumed diminishing returns involved in R&D activities, the reduction in marginal cost yielded by new R&D activities must necessarily be bounded at the margin. That is, the Inada condition  $\lim_{X \rightarrow 0} f'[X] = \infty$  exhibited by the R&D technology in the AJ model is difficult to defend. Finally, the quadratic R&D cost function yields a reduction in marginal cost that is proportional to the difference between the choke price and the initial marginal cost (D'Aspremont and Jacquemin (1988), page 1114). This result means that the higher is the initial marginal cost, the lower will be the marginal cost reduction. In particular, when the initial marginal cost is so high to be equal to the choke price, the marginal cost reduction will be zero, a result that is clearly unsatisfactory.

$$(3) \quad c_i^1 = c^0 - f[x_i + \beta \sum_{j \neq i} x_j].$$

Let  $q_i$  be the output of firm  $i$ . The profit obtained by firm  $i$  in the production stage and under the strategy profile  $(q_1, \dots, q_i, \dots, q_n)$  is

$$\varphi_i[q_1, \dots, q_i, \dots, q_n] = q_i(a - b(q_1 + \dots + q_i + \dots + q_n) - c_i^1).$$

In the production stage, firm  $i$  solves the following profit maximization problem:

$$(4) \quad \max_{q_i} \varphi_i[q_1, \dots, q_i, \dots, q_n].$$

$(i = 1, \dots, n)$ .

If the post-innovation cost of firm  $i$  is high, it will not be able to compete with the other firms and will choose not to produce, with the ensuing consequence that it makes zero profit in the production stage. On the other hand, if its post-innovation marginal cost is not too high, firm  $i$  will be able to produce a positive level of output and earn positive profits in the production stage. Thus, the first-order condition that characterizes the best response of firm  $i$  to  $(q_j)_{j \neq i}$  is

$$(5) \quad a - b(q_1 + \dots + q_i + \dots + q_n) - c_i^1 - bq_i \leq 0,$$

$(i = 1, \dots, n),$

with equality holding if  $q_i > 0$ . The first-order condition (5) is also sufficient because  $\varphi_i[q_1, \dots, q_i, \dots, q_n]$  is strictly concave in  $q_i$ .

The  $n$  first-order conditions (5) can be used to compute the equilibrium output of each firm, as a function of its post-innovation marginal cost. If we let  $q_i[c_1^1, \dots, c_n^1]$  denote the equilibrium output of firm  $i$  in the production stage, then the profit it earns in this stage is given by

$$(6) \quad \pi_i[c_1^1, \dots, c_n^1] = \varphi_i[q_1[c_1^1, \dots, c_n^1], \dots, q_i[c_1^1, \dots, c_n^1], \dots, q_n[c_1^1, \dots, c_n^1]].$$

Note that  $\pi_i[c_1^1, \dots, c_n^1] \geq 0$ , with strict inequality holding if  $c_i^1$  is not too high.

Let  $I$  denote the subset of firms that produce a positive level of output and a fortiori earns a positive level of profit in the production stage. For each  $i \in I$ , the first-order condition (5) holds with equality, and summing these first-order conditions over  $i \in I$ , we obtain

$$(7) \quad |I|a - |I|bQ - \sum_{i \in I} c_i^1 - bQ \leq 0,$$

where  $Q = q_1 + \dots + q_i + \dots + q_n$  denotes the industry output, and  $|I|$  denotes the number of elements in  $I$ , ie., the number of firms that produce a positive level of output.

It follows from (7) that

$$(8) \quad Q = \frac{|I|a - \sum_{i \in I} c_i^1}{(|I|+1)b}.$$

Using (8), we obtain the following expression for the equilibrium market price

$$(9) \quad p = a - bQ = \frac{a + \sum_{i \in I} c_i^1}{|I|+1}.$$

Using (8) in (5), we obtain the following expression for the output of a firm, say  $i$ , which produces a positive level of output in the production stage:

$$(10) \quad q_i = \frac{a - |I|c_i^1 + \sum_{j \in I, j \neq i} c_j^1}{(|I|+1)b},$$

( $i \in I$ ).

The profit made by firm  $i$  in the post-innovation stage is then given by

$$(11) \quad \pi_i[c_1^1, \dots, c_n^1] = q_i(p - c_i^1) \\ = \begin{cases} \frac{(a - |I|c_i^1 + \sum_{j \in I, j \neq i} c_j^1)^2}{(|I|+1)^2 b}, & \text{if } i \in I \\ 0, & \text{otherwise.} \end{cases}$$

## 2.4 The equilibrium in the innovation stage

Let  $x_j, j = 1, \dots, n$ , denote the own R&D expenditure made by firm  $j$  in the first stage. Because the profit made by a firm in the production stage is bounded above, it is not optimal for any firm to spend a large amount of its own resources on R&D. Thus, we shall assume that the own R&D made by each firm is constrained to belong to a closed bounded interval, say  $0 \leq x_j \leq K, j = 1, \dots, n$ , where  $K$  is a finite positive number. Given the list  $(x_1, \dots, x_n)$  of own R&D expenditures, the profit – net of R&D costs – earned by firm  $i$  over the two stages of the game is given by

$$(12) \quad \phi_i[x_1, \dots, x_i, \dots, x_n] = -x_i + \pi_i[c_1^1, \dots, c_i^1, \dots, c_n^1] = \\ -x_i + \pi_i[c^0 - f[x_1 + \beta \sum_{j \neq 1} x_j], \dots, c^0 - f[x_i + \beta \sum_{j \neq i} x_j], \dots, c^0 - f[x_n + \beta \sum_{j \neq n} x_j]].$$

The first-order condition that characterizes the best response of firm  $i$  to  $(x_j)_{j \neq i}$  is

$$(13) \quad \frac{\partial \phi_i[x_1, \dots, x_i, \dots, x_n]}{\partial x_i} = -1 - \frac{\partial \pi_i[c_1^1, \dots, c_i^1, \dots, c_n^1]}{\partial c_i^1} f'[x_i + \beta \sum_{j \neq i} x_j] \\ - \beta \sum_{j \neq i} \frac{\partial \pi_i[c_1^1, \dots, c_i^1, \dots, c_n^1]}{\partial c_j^1} f'[x_j + \beta \sum_{j' \neq j} x_{j'}] \leq 0,$$

with equality holding when  $x_i > 0$ .

Note that it is not optimal for a firm to spend a positive amount of its own resources on R&D and then chooses not to produce a positive level of output in the production stage. Furthermore, a firm might find it profitable to produce a positive level of output even when it did not spend any of its own resources on R&D in the first stage. Under this scenario, it takes advantage of the spillovers from the R&D activities of all the firms which choose to spend a positive amount of their own resources on their R&D activities to lower its own marginal cost.

When the own R&D expenditure of firm  $i$  is positive, the following second-order condition must also be satisfied:

$$(14) \quad \frac{\partial^2 \phi_i[x_1, \dots, x_i, \dots, x_n]}{\partial x_i^2} < 0.$$

Let  $\zeta_i[x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$  denote firm  $i$ 's best response – presumed to be unique – to  $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ ,  $i = 1, \dots, n$ . It is simple to show that the map

$$\zeta_i: (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \rightarrow \zeta_i[x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$$

is continuous. Thus, the map

$$(x_1, \dots, x_i, \dots, x_n) \rightarrow (\zeta_i[x_1, \dots, x_i, \dots, x_n])_{i=1}^n$$

is a continuous map from the convex compact subset  $[0, K]^n$  of the  $n$ -dimensional Euclidean space into itself, and thus, according to the Brouwer's fixed point theorem, will have a fixed point. The fixed point, which we presume to be unique and denote by  $(x_1[n, \beta], \dots, x_n[n, \beta])$ , represents the equilibrium list of own R&D expenditures made by the  $n$  firms. Note that the argument just presented establishes the existence of an equilibrium for any number of firms. Furthermore, for an arbitrary number of firms in the market, some firms might make zero net profit (net of own R&D cost). In particular, when the number of firms is large, the equilibrium might involve some firms being not active in the production stage, which necessarily means that these firms do not expend any of their own resources on R&D, and thus can be dropped from the game without affecting the equilibrium generated by the remaining firms. In other words, these firms will choose not to enter the market.

For an equilibrium under which each firm spends a positive amount of its own resources on R&D, (12) takes on the following more specific form:

$$(15) \quad \phi_i[x_1, \dots, x_i, \dots, x_n] = -x_i + \frac{1}{b(n+1)^2} \left( \frac{a - n(c^0 - f[x_i + \beta \sum_{j \neq i} x_j])}{+ \sum_{j \neq i} (c^0 - f[x_j + \beta(x_i + \sum_{j' \neq i, j' \neq j} x_{j'}))]} \right)^2.$$

Furthermore, the first-order condition (13) becomes

(16)

$$\frac{\partial \phi_i[x_1, \dots, x_i, \dots, x_n]}{\partial x_i} = -1 + \frac{2}{b(n+1)^2} \left( \frac{a - n(c^0 - f[x_i + \beta \sum_{j \neq i} x_j])}{+ \sum_{j \neq i} (c^0 - f[x_j + \beta(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})])} \right) \times \left( \frac{nf'[x_i + \beta \sum_{j \neq i} x_j]}{-\beta \sum_{j \neq i} f'[x_j + \beta(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})]} \right) = 0,$$

( $i = 1, \dots, n$ ),

and the second-order condition (14) becomes

$$(17) \quad \frac{\partial^2 \phi_i[x_1, \dots, x_i, \dots, x_n]}{\partial x_i^2} = \left( \frac{a - n(c^0 - f[x_i + \beta \sum_{j \neq i} x_j])}{+ \sum_{j \neq i} (c^0 - f[x_j + \beta(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})])} \right) \times \left( \frac{nf''[x_i + \beta \sum_{j \neq i} x_j]}{-\beta^2 \sum_{j \neq i} f''[x_j + \beta(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})]} \right) + \left( \frac{nf'[x_i + \beta \sum_{j \neq i} x_j]}{-\beta \sum_{j \neq i} f'[x_j + \beta(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})]} \right)^2 < 0.$$

Note that if each firm makes a positive own expenditure on R&D, i.e., if  $x_i[n, \beta] > 0$  for all  $i = 1, \dots, n$ , then the  $n$  first-order conditions in (16) are symmetric, and it is necessary that all the own R&D expenditures are the same, i.e.,  $x_1[n, \beta] = \dots = x_n[n, \beta] = x[n, \beta]$ , where we have used  $x[n, \beta]$  to denote their common own R&D expenditure. In this case, the equilibrium is a symmetric equilibrium under which each firm spends a positive amount of its own resources on R&D, and we refer to such an equilibrium as a *symmetric equilibrium with innovation*. It might also happen that in equilibrium some firms incur a positive level of own R&D cost while some other firms choose not to do so. In this case, the equilibrium is an *asymmetric equilibrium*, with one proper subset of the firms incurring a positive level of own R&D cost while all the firms outside this subset choose not to do so. Because the first-order conditions in (16) that characterize the own R&D expenditures of the firms

that choose to spend a positive amount of their own resources on R&D are symmetric, the own R&D expenditures of these firms must be the same. Finally, it might happen that in equilibrium none of the firms chooses to spend any of its own resources on R&D. In this case, the equilibrium is a symmetric equilibrium without innovation.

For a symmetric equilibrium with innovation, the effective R&D expenditure of each firm is given by

$$X_1[n, \beta] = \dots = X_n[n, \beta] = X[n, \beta] = x[n, \beta](1 + (n - 1)\beta),$$

where  $X[n, \beta]$  denote the firms' common effective expenditure. Furthermore, the first-order condition (16) is then reduced to

$$(18) \quad -1 + \frac{2(n-(n-1)\beta)}{b(n+1)^2} (a - c^0 + f[X[n, \beta]])f'[X[n, \beta]] = 0,$$

and the second-order condition (17) is reduced to

$$(19) \quad \frac{n-(n-1)\beta^2}{n-(n-1)\beta} (a - c^0 + f[X[n, \beta]])f''[X[n, \beta]] + (f'[X[n, \beta]])^2 < 0.$$

Because

$$\frac{2(n-(n-1)\beta)}{b(n+1)^2} (a - c^0 + f[X[n, \beta]])f'[X[n, \beta]] < 1$$

when  $n$  is large, the first-order condition (18) will fail to hold and a fortiori no symmetric equilibrium with innovation will exist.

For the case  $n = 1$ , the first-order condition (18) is reduced to

$$(20) \quad -1 + \frac{1}{2b} (a - c^0 + f[X[1, \beta]])f'[X[1, \beta]],$$

which is the first-order condition for maximizing (15). As for the second-order condition (19), it is reduced to

$$(21) \quad (a - c^0 + f[X[1, \beta]])f''[X[1, \beta]] + (f'[X[1, \beta]])^2 < 0,$$

which is the second-order condition for maximizing (15), namely the second-order condition for monopoly profit maximization.

We note in passing that if the curve  $X \rightarrow (a - c^0 + f[X])f'[X]$ ,  $X \geq 0$ , is downward-sloping, a main assumption in the model of Kamien et al. (1992), then the solution of the first-order condition (18) is unique, and a symmetric equilibrium with innovation, if it exists, is necessarily unique. If more than one value of  $X$  satisfies the first-order condition (18), then the second-order condition (19) must be used to eliminate the inappropriate value of  $X$  that satisfies this first-order condition. Another possibility is that there might be a value of  $X$  that satisfies both the first order condition (18) and the second-order condition (19), but such a value of  $X$  leads to a negative net profit for each firm when we set  $x_1 = \dots = x_n = \frac{X}{1+(n-1)\beta}$  in (15). Under such a scenario, there is no symmetric equilibrium under which  $X$  constitutes the effective R&D expenditure of each firm in the market.

The product price under a symmetric equilibrium with innovation when there are  $n$  firms in the market is given by

$$(22) \quad p[n, \beta] = \frac{a+n(c^0-f[X[n, \beta]])}{n+1}.$$

The net profit made by a firm under a symmetric equilibrium with innovation, when there are  $n$  firms in the market, is given by

$$(23) \quad -\frac{X[n, \beta]}{1+(n-1)\beta} + \frac{(a-(c^0-f[X[n, \beta]]))^2}{b(n+1)^2}.$$

*An example:* Suppose that the R&D production function is given by (2). For a symmetric equilibrium with innovation, the first-order condition (18) becomes

$$(24) \quad -1 + \frac{2(n-(n-1)\beta)}{b(n+1)^2} (a - c^0 e^{-\gamma X}) \gamma c^0 e^{-\gamma X} = 0.$$

The slope of the curve

$$(25) \quad \phi: X \rightarrow \phi[X] = -1 + \frac{2(n-(n-1)\beta)}{b(n+1)^2} (a - c^0 e^{-\gamma X}) \gamma c^0 e^{-\gamma X}, X \geq 0,$$

is

$$(26) \quad \phi'[X] = \frac{2(n-(n-1)\beta)\gamma^2 c^0}{b(n+1)^2} e^{-2\gamma X} (2c^0 - ae^{\gamma X}).$$

Observe that if  $c^0 \leq \frac{a}{2}$ , then  $\phi'[X] < 0$  is negative for all  $X > 0$ , and the curve  $\phi: X \rightarrow \phi[X]$  is downward-sloping. In this case, the solution of the first-order condition (18), if it exists, is unique. On the other hand, if  $c^0 > \frac{a}{2}$ , then the curve  $\phi: X \rightarrow \phi[X]$  has the shape of an inverted U: rising at first, attaining its global maximum at  $X = \frac{1}{\gamma} \text{Log} \left[ \frac{2c^0}{a} \right]$ , and then strictly declining to  $-\infty$  when  $X \rightarrow \infty$ . For the case  $c^0 > \frac{a}{2}$ , if the curve  $\phi: X \rightarrow \phi[X]$  does not cross the horizontal axis on its rising part, then a symmetric equilibrium with innovation does not exist. On the other hand, if the curve  $\phi: X \rightarrow \phi[X]$  crosses the horizontal axis on its rising part, then on its declining part it crosses the horizontal axis again. In this case, there exist two values of  $X$  that satisfy the first-order condition (18). At the first crossing the expression on the left side of (19) is equal to  $(n-1)(1-\beta)\beta c^0 > 0$ , and this means that the second-order condition (19) is not satisfied at the first crossing. Thus, a symmetric equilibrium with innovation, if it exists, must occur at the second crossing.

If we let  $Z = e^{-\gamma X}$ , then the first-order condition (24) becomes

$$(27) \quad -1 + \frac{2(n-(n-1)\beta)}{b(n+1)^2} (a - c^0 Z) \gamma c^0 Z = 0,$$

which is a quadratic equation in  $Z$ . The two roots of (27) are of the same sign, and are given by

$$(28) \quad \left\{ Z \rightarrow \frac{a\eta c^0 - \sqrt{\eta(-2b(1+n)^2 + a^2\eta)(c^0)^2}}{2\eta(c^0)^2}, Z \rightarrow \frac{a\eta c^0 + \sqrt{\eta(-2b(1+n)^2 + a^2\eta)(c^0)^2}}{2\eta(c^0)^2} \right\},$$

where we have let  $\eta = (n(1-\beta) + \beta)\gamma$ .

As can be seen from (28), the second root is positive. Hence the first root is also positive. If the second root is greater than 1, then it must be rejected because  $Z = e^{-\gamma X} < 1$ . On the other hand, if the second root is less than 1, then the first root is also less than 1. Furthermore, if the curve  $\phi: X \rightarrow \phi[X]$  crosses the horizontal axis

twice, then the first-order condition (24) has two roots, and the larger root corresponds to the second crossing, and this means that the smaller root of (27) is the correct value of  $Z$  for the effective R&D expenditure under the symmetric equilibrium with innovation. Thus, the effective R&D expenditure of a firm under a symmetric equilibrium with innovation is given by

$$(29) \quad X[n, \beta] = -\frac{1}{\gamma} \text{Log}[Z[n, \beta]],$$

where we have let

$$(30) \quad Z[n, \beta] = \frac{1}{2c^0} \left( a - \sqrt{\frac{a^2 \gamma (n(1-\beta) + \beta) - 2b(1+n)^2}{\gamma (n(1-\beta) + \beta)}} \right) < 1.$$

## 2.5 Entry

To fix ideas about the entry process, we shall assume that a firm will enter the market only if it can earn positive net profits. For any value of  $\beta$ , there are two possible scenarios to consider. Under the first scenario, there is a positive integer  $m$ , such that (i) when there are  $m$  firms in the market, each firm earns positive net profit in equilibrium, and (ii) when there are more than  $m$  firms in the market, at least one firm earns zero net profit in equilibrium. Because we assume that a firm only enters the market if it earn positive net profit, exactly  $m$  firms will enter the market under this scenario, and  $m$  then represents the equilibrium number of firms. Under the second scenario, for any positive integer  $m$ , there exists a positive integer  $n \geq m$ , such that under the equilibrium with  $n$  firms in the market each firm earns positive net profit. In this case, the entry process goes on indefinitely, and in the limit, the equilibrium number of firms is infinite. In what follows, we shall denote by  $n[\beta]$  the equilibrium number of firms. Under the first scenario,  $n[\beta] < \infty$ , while under the second scenario,  $n[\beta] = \infty$ .

## 2.6 The properties of the equilibrium

Proposition 1: *Under any equilibrium, the number of firms that expend a positive amount of their own resources on R&D is finite.*

PROOF: The proof is by *reductio ad absurdum*. Suppose that there exists an equilibrium under which an infinite number of firms choose to expend a positive amount of their own resources on R&D. The competition among these firms will drive the product price down to their post-innovation marginal cost, and each of them will earn zero profit in the production stage. The profit net of own R&D cost of each of these firms will then be negative, and this cannot occur in equilibrium. ■

In what follows, we denote by  $n^+[\beta]$  the number of firms in the market that choose to spend a positive amount of their own resources on R&D under the equilibrium with  $\beta$  as the degree of spillovers. Note that  $0 \leq n^+[\beta] \leq n[\beta]$ . Without any loss of generality, we can assume that the first  $n^+[\beta]$  firms are exactly the firms that choose to expend a positive amount of their own resources on R&D and that the firms which choose not to expend any of their own resources on R&D are the last  $(n[\beta] - n^+[\beta])$  firms.

Also, for any  $\beta$ , let  $\underline{n}[\beta]$  denote the critical number of firms such that for any positive integer  $n \leq \underline{n}[\beta]$ , there exists a symmetric equilibrium with innovation when  $n$  firms enter the market, but no symmetric equilibrium with innovation when there are  $\underline{n}[\beta] + 1$  firms in the market. It is clear that  $\underline{n}[\beta] \leq n[\beta]$ . Note that when  $\underline{n}[\beta] = n[\beta]$ , the equilibrium is a symmetric equilibrium with innovation under which exactly  $\underline{n}[\beta]$  firms will enter the market.

When the degree of spillovers is low, a firm cannot rely on the spillovers generated by the other firms to lower its own marginal cost. To obtain any desired reduction in marginal cost, a firm must bear most of the costs needed to run its own research lab, and this discourages entry. Furthermore, once a firm has entered the market, it must

expend a substantial amount of its own resources to generate a given level of cost reduction. Hence when the degree of spillovers is low, we can expect a small number of firms to enter the market, and once a firm has entered the market, it will expend a positive amount of its own resource on R&D. These intuitive results are confirmed in Proposition 2.

*Proposition 2: When the degree of spillovers is low and the initial common marginal cost is close to the choke price, the number of firms that enter the market is  $\underline{n}[\beta]$ , and the equilibrium is a symmetric equilibrium with innovation.*

PROOF: Under the symmetric equilibrium with innovation that prevails after  $\underline{n}[\beta]$  firms have entered the market, the product price, according to (20), is given by

$$(31) \quad p[\underline{n}[\beta], \beta] = \frac{a+n(c^0-f[X[\underline{n}[\beta], \beta])}{n+1}.$$

Observe that  $p[\underline{n}[\beta], \beta]$  will much lower than  $c^0$  if  $c^0$  is close to the choke price  $a$ . Now if the first  $\underline{n}[\beta]$  firms continue to use the strategies associated with the symmetric equilibrium with  $\underline{n}[\beta]$  firms, then firm  $\underline{n}[\beta] + 1$ , which does not spend any of its own resources on R&D, will not manage to lower its cost below  $p[\underline{n}[\beta], \beta]$  because of the low degree of spillovers, and thus will not be able to produce any positive output in order to earn positive profits. Thus, no more entry will take place after  $\underline{n}[\beta]$  firms have entered the market. ■

The following lemma asserts that at some stage during the entry process, if an equilibrium under which none of the firms chooses to expend its own resource on R&D is reached, then the entry process will continue indefinitely, and in the limit the equilibrium is the competitive equilibrium without innovation.

*Lemma 1: For any value of  $\beta$ , if there exists a positive integer  $n$  such that under the equilibrium with  $\beta$  as the degree of spillovers and  $n$  as the number of firms in the market no firm chooses to expend its own resource on R&D, then the entry process*

*continues indefinitely, and in the limit, the resulting equilibrium is the competitive equilibrium without innovation.*

PROOF: Suppose that  $n$  firms have already entered the market. Let us consider the problem faced by firm  $n + 1$ , which is contemplating entering the market. Let us imagine that the marginal cost of firm  $n + 1$  in the production stage is equal to the product's choke price. Under such a scenario, this firm will choose not to produce any positive level of output, and this means that the equilibrium in the production stage with firms  $1, \dots, n, n + 1$  in the market is identical with the equilibrium with  $n$  firms in the market. Now let us lower the marginal cost of firm  $n + 1$  from the level of the product's choke price to  $c^0$ , the initial common marginal cost, while maintaining the marginal costs of firm 1 to firm  $n$  at  $c^0$ . During the process, the profit in the production stage of each of the firms  $1, \dots, n$  will be falling.<sup>7</sup> When the marginal cost of firm  $n + 1$  descends to  $c^0$ , each of the  $n + 1$  firms will produce the same level of output at the same marginal cost  $c^0$ , and earns the same level of profit in the production stage. Next, note that according to the hypothesis of the lemma, none of the firms finds it profitable to expend its own resource on R&D when the  $n - 1$  remaining firms choose not to incur any R&D cost on their own. Hence when one more firm enters the market and  $n$  of them choose not to expend any of their own resources on R&D, the remaining firm, which faces more competition when there are  $n + 1$  firms in the market than when there are  $n$  firms in the market, will not find it profitable, either, to expend its own resource on R&D. We have just demonstrated that under the equilibrium with  $n + 1$  firms in the market none of the firms chooses to expend its own resource on R&D.

The argument just presented can be repeated *ad infinitum* to show that the entry process will continue indefinitely, and in the limit the equilibrium is the competitive equilibrium without innovation. ■

---

<sup>7</sup> It is well known in Cournot oligopoly theory with linear demand curve and constant marginal costs that a fall in the marginal cost of a firm improves its profitability, but reduces the profit made by each of the other firms.

Lemma 2: *Suppose that for some  $\beta$ , there exists a positive integer  $n$  such that under the equilibrium with  $\beta$  as the degree of spillovers and with  $n$  as the number of firms in the market none of the firms chooses to expend any of its own resource on R&D. Then for any  $\beta'$ ,  $\beta < \beta' \leq 1$ , there exists a positive integer  $n' \leq n$  such that under the equilibrium with  $\beta'$  as the degree of spillovers and with  $n'$  as the number of firms in the market none of the firms chooses to expend any of its own resource on R&D.*

PROOF: According to the hypotheses of Lemma 1, if  $\beta$  is the degree of spillovers then none of the  $n$  firms will expend a positive amount of its own resource on R&D when the remaining  $n - 1$  firms choose not to spend any positive amount of their own resources on R&D. This statement is still true when the degree of spillovers is  $\beta' > \beta$ , and when the number of firms in the market is still  $n$ . ■

Lemma 3: *If  $\beta$  is close to 1, then there exists a positive integer  $n$  such that under the equilibrium with  $\beta$  as the degree of spillovers and with  $n$  as the number of firms in the market none of the firms chooses to expend any of its own resource on R&D.*

PROOF: The proof of Lemma 3 requires some limiting arguments, and is given in Annex A.

When there are full spillovers ( $\beta = 1$ ), the entire own R&D expenditure made by a firm flows freely to its rivals, and any rival firm – at no cost of its own – can obtain the same cost reduction obtained by the former firm. The full R&D spillovers destroy the incentive for any firm to spend its own resources on R&D. Indeed, if in equilibrium a firm chooses to incur a positive level of own R&D cost, then an infinite number of firms will enter the market and operate at the same marginal cost as the firm that spends a positive amount of its own resource on R&D. Under this scenario, there will be perfect competition in the post-innovation stage, with the ensuing consequence that all the firms in the market will earn zero profit (gross of own R&D cost) in the production stage. Thus, when there are full spillovers the equilibrium market structure is that of perfect competition without innovation. More generally,

when the degree of spillovers is high, a firm that chooses to expend a positive amount of its own resource on R&D bears the entire burden of its own R&D cost, but most of the fruit of its R&D activities flow freely to its rivals, and this destroys the incentive for a firm to spend its own resource on R&D, as asserted by the following proposition.

*Proposition 3: If  $\beta$  is close to 1, then an infinite number of firms will enter the market, and a firm – once it has entered the market – will choose not to spend any of its own resources on R&D. That is, when the degree of spillovers is high, there will be no innovation, and the resulting equilibrium market structure is that of perfect competition without innovation.*

PROOF: To prove Proposition 3, invoke Lemma 3 and then Lemma 1. ■

Proposition 2 asserts the existence of a symmetric equilibrium with innovation when the degree of spillovers is low, while Proposition 3 asserts that perfect competition without innovation will prevail when the degree of spillovers is high. Proposition 4 asserts the existence of an asymmetric equilibrium – an equilibrium under which some firms choose to expend a positive amount of their own resources on R&D while other firms choose not to – when the degree of spillovers take on the intermediate values. Proposition 4 asserts that when the degree of spillovers is low, a firm that wishes to enter the market cannot rely on the spillovers to lower its own marginal cost and must bear most of the burden required to lower its own marginal cost. This factor reduces the incentive for entry, and the equilibrium is a symmetric equilibrium with innovation under which a small number of firms choose to enter the market. When the degree of spillovers has risen sufficiently, the spillovers allow more firms to enter the market, and some of the entrants can free ride on the R&D activities of other entrants: the equilibrium is then an asymmetric equilibrium. When the degree of spillovers reaches a high level, there is no incentive for any firm to expend its own

resources on R&D, and the resulting equilibrium is the competitive equilibrium without innovation.

*Proposition 4: Suppose that pre-innovation common marginal cost is high. There exist two values of  $\beta$ , say  $\underline{\beta}$  and  $\bar{\beta}$ , with  $0 < \underline{\beta} < \bar{\beta} < 1$ , which have the following properties:*

- (i) *For all  $0 < \beta < \underline{\beta}$ , the equilibrium with  $\beta$  as the degree of spillovers is a symmetric equilibrium with innovation.*
- (ii) *For all  $\bar{\beta} < \beta \leq 1$ , the equilibrium with  $\beta$  as the degree of spillovers is the competitive equilibrium without innovation.*
- (iii) *There exists a value of  $\beta \in (\underline{\beta}, \bar{\beta})$ , such that the equilibrium with  $\beta$  as the degree of spillovers is an asymmetric equilibrium.*

PROOF: The proof of Proposition 4 is technical and is given in Annex B.

Lemma 4 deals with the effect of low degrees of spillovers on consumers' surplus, producers' surplus, and social welfare. In proving (iii) of Lemma 4, we assume that the R&D technology is given by (2).

*Lemma 4: When  $\beta$  rises in a small right neighborhood of 0, (i) consumers' surplus fall (ii) producers' surplus rises; and (iii) social welfare rises.*

PROOF: The proof of Lemma 4 involves the computations of various derivatives, and is quite technical. It is relegated to Annex C.

The following proposition describes how the degree of spillovers affects social welfare.

*Proposition 5: Suppose that the functional form of the R&D production function is given by (2). As the degree of spillovers rises from 0 to 1, social welfare first rises*

*with  $\beta$ , reaches a maximum, and then declines to the lowest possible level, which is the social welfare associated with the competitive equilibrium.*

PROOF: Proposition 5 follows immediately from Proposition 2, Proposition 3, and Lemma 4. ■

The economic contents of Proposition 5 embody the tension between allocative efficiency and dynamic efficiency. When the degree of spillovers is low, a firm can appropriate most of the fruit of its own R&D, and this encourages innovation. This incentive is reduced when the degree of spillovers rises, and each firm spends less of its own resources on R&D. On the other hand, the rise in the degree of spillovers encourages entries. Social welfare rises with the degree of spillovers when it rises slightly above zero because the allocative efficiency effect dominates the dynamic efficiency effect. After the degree of spillovers has reached a sufficiently high level, the spillovers discourage firms from spending their own resources on R&D, with the ensuing consequence that the equilibrium is the competitive equilibrium without innovation: the rise in allocative efficiency generated by the high number of firms scarcely counteracts the dramatic loss in dynamic efficiency. For intermediate values for the degree of spillovers, the initial rise in social welfare is reversed when the gain in allocative efficiency cannot offset the loss in dynamic efficiency.

## **2.7 A numerical simulation**

In the numerical simulation, the R&D technology is assumed to have the functional form represented by (2). To compute the equilibrium number of firms for the two-stage game, we proceed as follows. We begin with the monopoly case, and then successively raise the number of firms by 1 each time. At each step, use (29) to compute the effective R&D expenditure of the symmetric equilibrium with innovation for the current number of firms. If the symmetric equilibrium with innovation when there are  $n$  firms cannot deter entry, then raise the number of firms

to  $n + 1$ , and then try to compute the symmetric equilibrium with innovation when  $n + 1$  firms are in the market. If there is no symmetric equilibrium with innovation for the case of  $n + 1$  firms – either because there is no positive value of  $X$  that satisfies the first-order condition (24) or because there exists a positive value of  $X$  that satisfies the first-order condition (24), but using it will result in a negative profit net of R&D cost for each of the  $n + 1$  firms – then the equilibrium must be an asymmetric equilibrium under which some firms choose to incur a positive level of own R&D cost, while others choose not to do so.

At any step of the procedure just described, entry will occur if it is profitable for new firms to enter the market. The first time it is not profitable for a new firm to enter the market, we have found the equilibrium number of firms. The following table summarizes the results of the simulations<sup>8</sup> we carried out for various values of  $\beta$ .

---

<sup>8</sup> The results of the simulation are obtained with the help of a large number of *Mathematica* programs.

**Table 1.****The equilibria for various values of  $\beta$  (Parameter values:  $a = 3, b = 0.25, c^0 = 2.75, \gamma = 2.75$ )**

|               | $n[\beta]$ | $n^+[\beta]$ | $x^+[\beta]$ | $c_i^1, i \leq n^+[\beta]$ | $c_i^1, i > n^+[\beta]$ | $p[\beta]$ | $PS[\beta]$ | $CS[\beta]$ | $SW[\beta]$ |
|---------------|------------|--------------|--------------|----------------------------|-------------------------|------------|-------------|-------------|-------------|
| $\beta = 0$   | 4          | 4            | 1.21         | 0.10                       | NA                      | 0.68       | 0.54        | 10.78       | 11.32       |
| $\beta = 0.1$ | 5          | 5            | 0.81         | 0.12                       | NA                      | 0.60       | 0.57        | 11.49       | 12.06       |
| $\beta = 0.2$ | 88         | 6            | 0.45         | 0.24                       | 0.63                    | 0.63       | 1.06        | 11.22       | 12.28       |
| $\beta = 0.3$ | $\infty$   | 6            | 0.30         | 0.35                       | 0.63                    | 0.63       | 0.02        | 11.25       | 11.27       |
| $\beta = 0.4$ | $\infty$   | 4            | 0.30         | 0.46                       | 0.74                    | 0.74       | 0.14        | 10.19       | 10.33       |
| $\beta = 0.5$ | $\infty$   | 3            | 0.28         | 0.60                       | 0.88                    | 0.88       | 0.10        | 8.96        | 9.06        |
| $\beta = 0.6$ | $\infty$   | 2            | 0.27         | 0.83                       | 1.12                    | 1.12       | 0.13        | 7.09        | 7.22        |
| $\beta = 0.7$ | $\infty$   | 1            | 0.29         | 1.23                       | 1.57                    | 1.57       | 0.16        | 4.09        | 4.25        |
| $\beta = 0.8$ | $\infty$   | 0            | NA           | NA                         | NA                      | 2.75       | 0           | 0.125       | 0.125       |
| $\beta = 0.9$ | $\infty$   | 0            | NA           | NA                         | NA                      | 2.75       | 0           | 0.125       | 0.125       |
| $\beta = 1$   | $\infty$   | 0            | NA           | NA                         | NA                      | 2.75       | 0           | 0.125       | 0.125       |

The first row of the table gives the equilibrium for the case of no spillovers ( $\beta = 0$ ). Under this equilibrium, 4 firms enter the market ( $n[\beta] = 4$ ), and each of them ( $n^+[\beta] = 4$ ) chooses to spend a positive amount ( $x^+[\beta] = 1.21$ ) of its own resource on R&D. The equilibrium is thus a symmetric equilibrium with innovation. The post-innovation marginal cost is  $c_i^1 = 0.10, i = 1, \dots, n[\beta]$ . The equilibrium product price is  $p[\beta] = 0.68$ . As for welfare, the producers' surplus (net of R&D cost) is  $PS[\beta] =$

0.54 and the consumer surplus is  $CS[\beta] = 10.78$ , which sum up to a level of social welfare given by  $SW[\beta] = 11.32$ .

The second line of the table gives the equilibrium for the case  $\beta = 0.1$ . The equilibrium is a symmetric equilibrium with innovation under which 5 firms enter the market. The amount of own resource that a firm spends on R&D is 0.81, which yields an effective R&D of  $0.81(1 + (5 - 1)(0.1)) = 0.81(1.4) = 1.134$ , which is less than 1.21, the effective R&D expenditure of each firm under the symmetric equilibrium with innovation for  $\beta = 0$ . The post-innovation cost is  $c_i^1 = 0.12, i = 1, \dots, n[\beta]$ . The equilibrium product price is 0.60, which is lower than the value it takes under the symmetric equilibrium with innovation for the case of no spillovers. The lower product price implies a higher level of consumers' surplus. The producers' surplus is also higher. The end result is a higher level of social welfare. The spillovers make it possible for one more firm to enter the market. Although the post-innovation marginal cost is higher than when there are no spillovers, the gain in welfare due to a higher number of firms more than offset the impact of the higher post-innovation marginal cost relatively to the case of no spillovers.

The third line gives the equilibrium for the case  $\beta = 0.2$ . The equilibrium number of firms is now 88, with 6 firms choosing to spend a positive amount of their own resources ( $x^+[\beta] = 0.45$ ) on R&D, while the remaining firms choose not to incur any own R&D cost. The equilibrium is asymmetric. The post-innovation marginal cost for the firms that spend a positive amount of their own resources on R&D is  $c_i^1 = 0.24, i = 1, \dots, n^+[\beta]$ . The post-innovation marginal cost for those who free ride on the R&D activities of other firms is  $c_i^1 = 0.63, i = n^+[\beta] + 1, \dots, n[\beta]$ . The equilibrium product price is 0.63, which is slightly higher than that under the equilibrium for the case  $\beta = 0.1$ , and this implies a slightly lower level of consumers' surplus ( $11.22 < 11.49$ ). The much lower own R&D cost of each firm helps to raise the producers' surplus substantially ( $1.06 > 0.57$ ). Social welfare is higher ( $12.28 > 12.06$ ) than the level attained for the case  $\beta = 0.1$ .

The fourth line gives the equilibrium for the case  $\beta = 0.3$ . The equilibrium number of firms is now infinite, with 6 firms choosing to spend a positive amount of their own resources ( $x^+[\beta] = 0.30$ ) on R&D, while the remaining firms choose not to incur any own R&D cost. The equilibrium is asymmetric. The post-innovation marginal cost for the firms that spend a positive amount of their own resources on R&D is  $c_i^1 = 0.35, i = 1, \dots, n^+[\beta]$ . The post-innovation marginal cost for those who free ride on the R&D activities of other firms is  $c_i^1 = 0.63, i = n^+[\beta] + 1, \dots, n[\beta]$ . The equilibrium product price is 0.63. The producers' surplus is 0.02 and the consumers' surplus is 11.25, which sum up to a social welfare level of 11.27. Note that social welfare has dropped when  $\beta$  rises from 0.2 to 0.3. The fall in social welfare is due to the lower own R&D expenditures made by the firms that choose to do so and the competition from the fringe made possible by the higher degree of spillovers.

The fifth line gives the equilibrium for the case  $\beta = 0.4$ . The equilibrium number of firms is infinite, with 4 firms choosing to spend a positive amount of their own resources ( $x^+[\beta] = 0.30$ ) on R&D, while the remaining firms choose not to incur any own R&D cost. The equilibrium is asymmetric. The post-innovation marginal cost for the firms that spend a positive amount of their own resources on R&D is  $c_i^1 = 0.46, i = 1, \dots, n^+[\beta]$ . The post-innovation marginal cost for those who free ride on the R&D activities of other firms is  $c_i^1 = 0.74, i \geq n^+[\beta] + 1$ . The equilibrium product price is 0.74. The producers' surplus is 0.14 and the consumers' surplus is 10.19, which sum up to a social welfare level of 10.33. Note that the producers' surplus rises when  $\beta$  rises from 0.3 to 0.4 because of the lower number of firms that incur a positive level of own R&D cost. Also, note that social welfare has dropped when  $\beta$  rises from 0.3 to 0.4. The fall in social welfare is due to the lower own R&D expenditures made by the firms that choose to do so and the competition from the fringe made possible by the higher degree of spillovers.

The sixth line gives the equilibrium for the case  $\beta = 0.5$ . An infinite number of firms enter the market, and among them 3 firms choose to spend a positive amount of their

own resources on R&D. The post-innovation marginal cost of the 3 dominant firms is 0.60, while the post-innovation marginal cost of the firms of the fringe is 0.88. Relatively to the case  $\beta = 0.5$ , the producers' surplus (0.10) is higher; the consumers' surplus is lower; and social welfare is lower.

The seventh line gives the equilibrium for the case  $\beta = 0.6$ . An infinite number of firms enter the market, and among them 2 firms choose to spend a positive amount of their own resources on R&D. The post-innovation marginal cost of the 2 dominant firms is 0.83, while the post-innovation marginal cost of the firms of the fringe is 1.12. Relatively to the case  $\beta = 0.6$ , the producers' surplus (0.13) is higher; the consumers' surplus is lower; and social welfare (7.22) is lower.

The eighth line gives the equilibrium for the case  $\beta = 0.7$ . An infinite number of firms enter the market, and among them only one firm chooses to spend a positive amount of their own resources on R&D. The post-innovation marginal cost of the dominant firm is 1.23, while the post-innovation marginal cost of the firms of the fringe is 1.57. Relatively to the case  $\beta = 0.7$ , the producers' surplus (0.13) is higher; the consumers' surplus is lower; and social welfare (4.25) is lower.

The last three lines of the table give the equilibria for the cases  $\beta = 0.8, 0.9$ , and  $1.0$ , respectively. An infinite number of firms enter the market, and the resulting equilibrium market structure is perfect competition without innovation. Under perfect competition, the market price is equal to the initial common marginal cost; the producers' surplus is 0; and the consumers' surplus is 0.125. Thus, high values for the degree of spillovers yield the lowest level of social welfare possible.

The numerical simulation indicates that social welfare is at its highest level at  $\beta = 0.2$ . The market structure is that of an asymmetric equilibrium under which some firms expend a positive amount of their own resources on R&D while many firms free ride on the R&D activities of others to lower their own marginal cost.

## 2.8 Conclusion

In this essay we have formulated a model of the AJ type to analyze the impact of R&D spillovers on entry and the resulting equilibrium market structure. We find that the degree of spillovers plays a fundamental role on the number of firms entering the market, their R&D activities, and social welfare. Our analysis suggests that social welfare is maximized at some intermediate degree of spillovers. The policy implication of this result is that neither complete protection of intellectual property right nor lax enforcement of patent laws is socially optimal. Uncertainty and risk are important factors in R&D, but they have been ignored in the literature. These factors merit more attention. A more complete modeling of the innovation process should include an examination of the major drivers influencing the degree of spillovers: distance between the innovators, property rights, and the extent of telecommunication networks.

## Annex A: The Proof of Lemma 3

We prove Lemma 3 by *reductio ad absurdum*. If Lemma 3 is not true, then we can find a sequence  $\beta_k, k = 0, 1, \dots$ , with  $\beta_k < 1, \beta_k \uparrow 1$ , such that for any  $k$  and any positive integer  $m$ , at least one firm will expend a positive amount of its own resource on R&D under the equilibrium with  $\beta_k$  as the degree of spillovers and  $m$  as the number of firms in the market. If the equilibrium number of firm when  $\beta_k$  is the degree of spillovers is finite, then a possible value for  $m$  is  $m_k = n[\beta_k]$ . When  $n[\beta_k] = \infty$ , a possible value for  $m$  is a positive integer  $m_k$  that satisfies the condition  $m_k > \frac{1}{1-\beta_k}$ .

According to (16), the following first-order condition characterizes the own R&D expenditure of a firm, say firm  $i$ , which expends a positive amount of its own resource on R&D under the equilibrium with  $\beta_k$  as the degree of spillovers and  $m_k$  as the number of firms in the market:

$$(A.1) \quad -1 + \frac{2}{b(m_k+1)^2} \left( \frac{a - m_k(c^0 - f[x_i + \beta_k \sum_{j \neq i} x_j])}{\sum_{j \neq i} (c^0 - f[x_j + \beta_k(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})])} \right) \times \\ \left( \frac{m_k f'[x_i + \beta_k \sum_{j \neq i} x_j]}{-\beta_k \sum_{j \neq i} f'[x_j + \beta_k(x_i + \sum_{j' \neq i, j' \neq j} x_{j'})]} \right) = 0.$$

Now because  $\beta_k \uparrow 1$  when  $k \rightarrow \infty$ , all of the firms, regardless of whether they choose to expend their own resources on R&D or choose to free ride on the R&D activities of others, will have the same post-innovation marginal cost in the limit. Hence the equilibrium number of firms, namely  $n[\beta_k]$  will be indefinitely large when  $k \rightarrow \infty$ , and this implies  $\lim_{k \rightarrow \infty} m_k = \infty$ . Furthermore, the total industry R&D expenditure must tend to 0 when  $k \rightarrow \infty$ . Indeed, if this is not the case, then the producers' surplus – the total industry profits in the production stage minus the total industry R&D expenditure – will be negative, and this cannot hold in equilibrium. Thus, in the limit, the first-order condition (A.1) becomes

$$(A.2) \quad -1 + \left( \lim_{k \rightarrow \infty} \frac{2}{b(m_k+1)^2} \right) (a - c^0) f'[0] = -1 < 0,$$

which is not consistent with (A.1) in the limit. ■

## Annex B: The Proof of Proposition 4

Let  $\underline{\beta}$  be the least upper bound of the degrees of spillovers  $\beta$ , such that for all  $0 < \beta' < \underline{\beta}$ , the equilibrium associated with  $\beta'$  is a symmetric equilibrium with innovation. The existence of  $\underline{\beta}$  follows directly from Proposition 2. Using Proposition 3, we can assert the existence of a value, say  $\bar{\beta} < 1$ , for the degree of spillovers, such that for all  $\bar{\beta} < \beta \leq 1$ , the equilibrium that prevails is perfect competition without innovation. To prove Proposition 4, we first establish a series of claims.

Claim 1: We have  $\underline{\beta} < \bar{\beta}$ .

PROOF: First, we claim that it is not possible to have  $\bar{\beta} < \underline{\beta}$ . Indeed, if this were the case, then for each value of  $\beta \in (\bar{\beta} < \underline{\beta})$ , the equilibrium associated with  $\beta$  is both a symmetric equilibrium with innovation under which each firm earns net positive profit and the perfectly competitive equilibrium without innovation, and this is absurd. Thus,  $\underline{\beta} \leq \bar{\beta}$ . ■

Claim 2:  $\underline{\beta} \neq \bar{\beta}$ .

PROOF: The claim is proved by *reductio ad absurdum*. Suppose that  $\underline{\beta} = \bar{\beta}$ . Using the definition of  $\underline{\beta}$ , we can find a symmetric equilibrium with innovation when the degree of spillovers is  $\underline{\beta} - \frac{1}{k}$  for large positive integers  $k$ . Such an equilibrium will converge to the equilibrium with  $\underline{\beta}$  as the degree of spillovers when  $k \rightarrow \infty$ . Furthermore, because the first-order condition (18) must be satisfied by each of these equilibria, in the limit, the equilibrium number of firms under the equilibrium with  $\underline{\beta}$  as the degree of spillovers must be finite, i.e.,  $n[\underline{\beta}] < \infty$ .

Using the definition of  $\bar{\beta}$ , we can find a sequence of degrees of spillovers  $(\beta_k)_{k=1}^{\infty}$ , with  $\beta_k \downarrow \bar{\beta}$ , such that for each  $k = 1, 2, \dots$ , the equilibrium with  $\beta_k$  as the degree of spillovers is a competitive equilibrium without innovation. In the limit when  $k \rightarrow \infty$ , these equilibria converge to a competitive equilibrium without innovation. Thus, the equilibrium number of firms when  $\bar{\beta}$  is the degree of spillovers is  $n[\bar{\beta}] = \infty$ . Thus, if  $\underline{\beta} = \bar{\beta}$ , then the equilibrium number of firms when the degree of spillovers is  $\underline{\beta}$  will be both finite and infinite, and this is absurd. ■

Claim 3: For some  $\epsilon > 0$  sufficiently small, there exists an asymmetric equilibrium with  $\beta \in (\underline{\beta}, \underline{\beta} + \epsilon)$ .

PROOF: First, note that there is no perfectly competitive equilibrium without innovation for each possible value for the degree of spillovers  $\beta \in (\underline{\beta}, \underline{\beta} + \epsilon)$ . Indeed, if this is not true, then we can find a sequence of degrees of spillovers decreasing to  $\underline{\beta}$ , such that the equilibrium associated with each of these degrees of spillovers is a competitive equilibrium without innovation, and this cannot be true by the argument used to establish Claim 2. Next, note that if there is no asymmetric equilibrium associated with some  $\beta \in (\underline{\beta}, \underline{\beta} + \epsilon)$ , then the equilibria associated with all  $\beta \in (\underline{\beta}, \underline{\beta} + \epsilon)$  must be symmetric equilibria with innovation, and this contradicts the fact that  $\underline{\beta}$  is the least upper bound of the values of  $\beta$  such that for all  $\beta' < \beta$ , the equilibrium with  $\beta'$  as the degree of spillovers is a symmetric equilibrium with innovation.

Together, the three claims constitute Proposition 4. ■

### Annex C: The Proof of Lemma 4

To prove (i) of Lemma 4, first note that the consumers' surplus under a symmetric equilibrium with innovation is given by

$$(C.1) \quad CS[n, \beta] = \frac{1}{2}(a - p[n, \beta])Q[n, \beta],$$

where, according to (9),

$$(C.2) \quad p[n[\beta], \beta] = \frac{a + n(c^0 - f[X[n[\beta], \beta]])}{n+1}$$

is the equilibrium product price, and

$$(C.3) \quad Q[n[\beta], \beta] = \frac{na - n(c^0 - f[X[n[\beta], \beta]])}{(n[\beta] + 1)b},$$

according to (8), is the equilibrium industry output.

Now the equilibrium number of firms when there are no spillovers is  $n[0]$ . As  $\beta$  rises slightly from 0, the equilibrium number of firms remains at the same level, i.e.,  $n[\beta] = n[0]$  when  $\beta$  is small. Next, note that for  $n[\beta] = n[0]$ , the first-order condition (18) shifts downward as  $\beta$  rises slightly from 0. Hence the equilibrium effective expenditure under the symmetric equilibrium with innovation falls – and this means that the equilibrium product price (C.2) rises – when  $\beta$  rises slightly from 0. The rise in the equilibrium product price implies a fall in the consumers' surplus, and (i) of Lemma 4 is proved.

To prove (ii) of Lemma 4, first, note that the producers' surplus under a symmetric equilibrium with innovation is given by

$$PS[n[\beta], \beta] = -\frac{n[\beta]X[n[\beta], \beta]}{1+(n[\beta]-1)\beta} + \frac{n[\beta]}{b(n[\beta]+1)^2} (a - c^0 + f[X[n[\beta], \beta]])^2,$$

which assumes the following form when the degree of spillovers is low:

$$(C.4) \quad PS[n[0], \beta] = -\frac{n[0]X[n[0], \beta]}{1+(n[0]-1)\beta} + \frac{n[0]}{b(n[0]+1)^2} (a - c^0 + f[X[n[0], \beta]])^2.$$

Differentiating (C.4) with respect to  $\beta$ , we obtain

$$(C.5) \quad \frac{1}{n[0]} \frac{\partial PS[n[0], \beta]}{\partial \beta} = -\frac{1}{1+(n[0]-1)\beta} \frac{\partial X[n[0], \beta]}{\partial \beta} + \frac{(n[0]-1)X[n[0], \beta]}{(1+(n[0]-1)\beta)^2} + \\ + \frac{2(a-c^0+f[X[n[0], \beta]])f'[X[n[0], \beta]]}{b(n[0]+1)^2} \frac{\partial X[n[0], \beta]}{\partial \beta}.$$

Multiply (C.5) with  $(n[0] - (n[0] - 1)\beta)$ , we obtain

$$(C.6) \quad \frac{(n[0]-(n[0]-1)\beta)}{n[0]} \frac{\partial PS[n[0], \beta]}{\partial \beta} = \\ -\frac{(n[0]-(n[0]-1)\beta)}{1+(n[0]-1)\beta} \frac{\partial X[n[0], \beta]}{\partial \beta} + \frac{(n[0]-(n[0]-1)\beta)(n[0]-1)X[n[0], \beta]}{(1+(n[0]-1)\beta)^2} + \\ + \frac{2(n[0]-(n[0]-1)\beta)(a-c^0+f[X[n[0], \beta]])f'[X[n[0], \beta]]}{b(n+1)^2} \frac{\partial X[n[0], \beta]}{\partial \beta}$$

$$= -\frac{(n[0]-(n[0]-1)\beta)}{1+(n[0]-1)\beta} \frac{\partial X[n[0],\beta]}{\partial \beta} + \frac{(n[0]-(n[0]-1)\beta)(n[0]-1)X[n,\beta]}{(1+(n[0]-1)\beta)^2} + \frac{\partial X[n[0],\beta]}{\partial \beta}.$$

Note that the third line in (C.6) has been obtained by using the first-order condition (18), which asserts that

$$\frac{2(n[0]-(n[0]-1)\beta)(a-c^0+f[X[n[0],\beta]])f'[X[n[0],\beta]]}{b(n[0]+1)^2} = 1.$$

Thus,

$$(C.7) \quad \frac{(n[0]-(n[0]-1)\beta)}{n[0]} \frac{\partial PS[n[0],\beta]}{\partial \beta} = \left(1 - \frac{(n[0]-(n[0]-1)\beta)}{1+(n[0]-1)\beta}\right) \frac{\partial X[n[0],\beta]}{\partial \beta} + \frac{(n[0]-(n[0]-1)\beta)(n[0]-1)X[n[0],\beta]}{(1+(n[0]-1)\beta)^2}$$

When  $\beta \rightarrow 0$ , (C.7) becomes

$$(C.8) \quad \frac{\partial PS[n[0],0]}{\partial \beta} = (1 - n[0]) \frac{\partial X[n[0],0]}{\partial \beta} + n[0](n[0] - 1)X[n[0],0] > 0.$$

Note that the strict inequality (C.8) follows from the result  $\frac{\partial X[n,\beta]}{\partial \beta} < 0$  and the fact that  $n > 1$ . That is,  $\frac{\partial PS[n[\beta],\beta]}{\partial \beta} > 0$  when  $\beta$  is small. Hence the producer surplus rises with  $\beta$  when  $\beta$  is small, and (ii) of Lemma 4 is proved.

To prove (iii) of Lemma 4, first, note that social welfare is given by

$$(C.9) \quad SW[n,\beta] = CS[n,\beta] + PS[n,\beta].$$

For low degrees of spillovers, we have<sup>9</sup>

---

<sup>9</sup> The symbolic calculations were carried out by *Mathematica*.

$$(C.10) \quad \frac{\partial SW[n[0],0]}{(n[0]-1)\partial\beta} = \frac{(n[0]-2)\left(a\sqrt{n[0]\gamma} + \sqrt{-2b(1+n[0])^2 + a^2n[0]\gamma}\right) - 4n[0]^2\sqrt{-2b(1+n[0])^2 + a^2n[0]\gamma} \operatorname{Log}[Z[n[0],0]]}{4n[0]\gamma\sqrt{-2b(1+n[0])^2 + a^2n[0]\gamma}}.$$

Because  $n[\beta] \geq 2$ , and because  $Z[n, \beta]$ , as given by (30), is less than 1, the right side of (C.10) is positive. Hence as  $\beta$  rises in a right neighbourhood of 0, social welfare also rises with  $\beta$ , and (iii) of Lemma 4 is proved. ■

## References

- Amir, R.; I. Evstigneev, and J. Wooders (2003)**, "Non-cooperative versus Cooperative R&D with Endogenous Spillover Rate," *Games and Economics Behaviour*, 42, 183-207.
- Atallah, G. (2000)**, "Vertical R&D Spillovers, Cooperation, Market Structure, and Innovation," *CIRANO Working Papers 2000s-54*, CIRANO.
- (2007)**, "Entry Deterrence through Fixed Cost Reducing R&D," *Rivista di Politica Economica*, 49-77.
- Cassiman, B. and R. Veugelers (2002)**, "Complementarity in the Innovation Strategy: Internal R&D, External Technology Acquisition, and Cooperation in R&D," *IESE Business School Working Paper No. 457*.
- Dasgupta, P. and J. Stiglitz (1980)**, "Industrial Structure and the Nature of Innovative Activity," *The Economic Journal*, 90, 266-293.
- D'Aspremont, C. and A. Jacquemin (1988)**, "Cooperative and Non-cooperative R&D in Duopoly with Spillovers," *The American Economics Review*, 78(5), 1133-37.
- De Bondt, R. (1996)**, "Spillovers and Innovative Activities," *International Journal of Industrial Organization*, 15(1), 1-28.
- De Bondt, R., P. Slaets, and B. Cassiman (1992)**, "The Degree of Spillovers and the Number of Rivals for Maximum Effective R&D," *International Journal of Industrial Organization*, 10, 35-54.
- Erkal, N. and D. Piccinin (2010)**, "Cooperative R&D under Uncertainty with Free Entry," *International Journal of Industrial Organization*, 28, 74-85.
- Kamien, M. I.; E. Muller, and I. Zang (1992)**, "Research Joint Ventures and R&D Cartels," *American Economic Review*, 82(5), 1293-306.
- Kamien, M. I. and I. Zang (1993)**, "Competing Research Joint Ventures," *Journal of Economics and Management Strategy*, 2(1), 23-40.
- Poyago-Theotoky, Joanna (1995)**, "Equilibrium and Optimal Size of a Research Joint Venture in an Oligopoly with Spillovers," *Journal of Industrial Economics*, Volume XLIII, 209-226.
- Ruff, L. E. (1969)**, "Research and Technological Progress in a Cournot Economy," *Journal of Economic Theory*, 1(4), 397-415.

- Spence, M. (1984)**, “Cost reduction, Competition and Industry Performance,” *Econometrica*, 52(1), 101–21.
- Suetens, S. (2005)**, “Cooperative and Non-cooperative R&D in Experimental Duopoly Markets,” *International Journal of Industrial Organization*, 23, 63-82.
- Suzumura, Kotaro (1992)**, “Cooperative and Non-cooperative R&D in an Oligopoly with Spillovers,” *American Economic Review*, 82, 1307-1320.
- Sveikauska, L. (2007)**, “R&D and Productivity Growth: A Review of the Literature,” *US Bureau of Labour Statistics*, Working Paper 408.
- Vives, X. (2008)**, “Innovation and Competitive Pressure,” *The Journal of Industrial Economics*, 56(3), 419-469.
- Yi, S. S. (1996)**, “The Welfare Effects of Cooperative R&D in Oligopoly with Spillovers,” *Review of Industrial Organization*, 11, 681-698.

### **3 The Search for New Drugs: A Theory of R&D in the Pharmaceutical Industry**

#### **3.1 Introduction**

In the economics of science and technology, economists distinguish between basic research and applied research. The objective of basic research is to increase our knowledge about natural phenomena, not to develop specific applications of this stock of knowledge for pecuniary gains. However, applied research – often called R&D or innovation in the economic literature – is driven by profits, and a firm carries out applied research in order to develop a new product or to find a more efficient production process. Another difference between basic research and applied research is that the discoveries of basic research are widely published and freely available to all researchers, while the results of R&D activities are jealously guarded secrets of the organizations that discover them in their efforts to appropriate all the benefits from these discoveries.

The economic analysis of basic research concentrates on the social value of the increase in the stock of fundamental knowledge. The social value of an increase in the stock of fundamental knowledge is difficult to measure, or even to predict. Because basic research takes place at the frontier of knowledge, the economic benefits of a program of basic research are highly uncertain. Furthermore, because scientific discoveries are difficult to establish or to defend, it is almost impossible to define the property rights of these discoveries. The difficulty involved in the definition of property rights in basic research implies that discoverers of new fundamental knowledge cannot appropriate all the commercial benefits of their research, and this form of market failure is the basis for the assertion that the level of investment in basic research is socially sub-optimal. This fact is reflected in the modest share of

basic research in the R&D budget of profit-driven firms. Yet for a program of basic research, the social value can greatly exceed its private value because advancements in fundamental knowledge constitute important inputs in many applied research programs and innovations.

Although R&D is profit-driven and mission-oriented, it shares one important characteristic with basic research: information externalities. The information generated by the R&D activities of a firm might help other firms in their search for a new product or a new process with lower costs. Revealing the results of its R&D activities to rival firms reduces a firm's competitive edge in its competition for profits and market share. This is the reason why firms keep the results of their R&D activities secret. However, rival firms can appropriate the fruits of the R&D activities of other firms through imitation, reverse engineering, or information spillovers. These factors tend to reduce the incentive of a firm to carry out R&D activities, and a solution to this problem is to offer some protection to a firm's innovation by granting it a patent.

Nordhaus (1969) was the first researcher to offer a rigorous analysis of the fundamental trade-off between static efficiency and dynamic efficiency in the design of patent policy. A patent gives its owner the intellectual property right to exploit the fruit of her discoveries, which encourages innovation, and thus benefits society. This is the source of dynamic efficiency behind the granting of a patent. However, offering protection to an innovation allows the holder of the patent to exercise monopoly power, which is the source of static inefficiency. The model of Nordhaus, *op cit.*, deals with the length of a patent, i.e., the number of years that the patent is in force.<sup>10</sup> It proposes complete protection of an innovation for a limited number of years, and can be taken as a simple description of the patent system in its original purpose. However, since the pioneering work of Mansfield (1961), researchers have found an overwhelming volume of evidence on the inability of patents to prevent imitation.

---

<sup>10</sup> In the US, the statutory life of a patent is about 17 years.

According to Mansfield (1984), about 60% of patented products were successfully imitated within four years of patenting. Nordhaus (1972) dealt with the issue of imitation by adding a second dimension – *patent breadth or patent width* – to his original model. Intuitively, the patent breadth captures the minimum degree of novelty relative to the product purportedly protected by the patent that a rival product must possess so that it can be judged as not infringing on the patent.

Although the concept of patent breadth is intuitively appealing, its meaning is vague. Nordhaus (1972) deals with process innovation, and the breadth of a patent is taken as the fraction of cost reduction not freely flowing to rival firms through spillovers. Klemperer (1990) analyzed patent breadth in the context of a model of spatially differentiated products. The distance in space between the patent holder's product and a rival product represents the transport cost that a consumer must pay if she buys the rival brand. The breadth of a patent is defined as the radius beyond which a rival product is considered as not to infringe upon the patent. Klemperer took the prize of the patent, i.e., the discounted profits enjoyed by the patent holder, as given, and determined the length and breadth of the patent that minimizes the social cost subject to the constraint that the patent holder earns the given prize. In the model of Gilbert and Shapiro (1990), patent breadth is simply defined as the flow rate of profit that a patent holder is allowed to earn while the patent is in force. The model is in reduced form, and the optimal length and breadth for the patent are found by maximizing discounted social welfare subject to the constraint that the patent holder is allowed to earn a certain level of profits. What this level of profits is was not specified by these researchers; neither was it explained in Klemperer's model. The varied definitions of patent breadth in the literature lead to an array of bewildering results. In some cases, the optimal patent has maximum length, but minimum breadth. In other cases, it is the opposite result. And in some other circumstances, the length-breadth mix makes no difference.<sup>11</sup> A consensus on the definition of patent breadth is yet to emerge.

---

<sup>11</sup> See Tandon (1982) for a discussion and a reconciliation of these contradictory results in a very simple model of the optimal length and breadth of a patent.

In the earlier literature on patent design, all efforts were concentrated on finding the optimal length and breadth of a patent. The role of information externalities was not considered. Indeed, one of the conditions for granting a patent is that the holder of the patent must provide a disclosure of all the technical information that enables another researcher to replicate all the results claimed by the patent. Furthermore, the analysis was carried out in the context of a single innovation, and the inherent uncertainties in R&D activities were ignored. An exception was Scotchmer and Green (1990), who formulated a simple dynamic economic model that incorporates patent breadth and the incentive of firms to keep secret their R&D results.

In the model of Scotchmer and Green, there are only two possible innovations: a weak innovation and a strong innovation. The weak innovation represents a slight improvement in the quality of the product in question, while the strong innovation represents a more pronounced improvement of the product. The weak innovation yields one extra unit of welfare in perpetuity, while the strong innovation yields two extra units of social welfare in perpetuity. R&D activities are driven by research expenditures, and the discovery time is modelled as the arrival of a Poisson process. In the model, the information externalities were modelled in the following manner. The weak innovation requires one Poisson hit, while the strong innovation requires two Poisson hits. If a firm has obtained a Poisson hit, then it only needs another Poisson hit to obtain strong innovation. If the technical information obtained by the firm that obtained the first Poisson hit is disclosed to the other firm, which has not made any Poisson hit, then the latter firm can benefit from the information disclosure and only needs one Poisson hit to obtain strong innovation. On the other hand, if the former firm does not patent its weak innovation, then the latter firm must make two Poisson hits to obtain strong innovation.

O'Donoghue et al. (1998) formulated a model in which firms sequentially improve the products of each other through time, and technological progress occurs when a non-infringing innovation displaces a patented product. In this model, innovations are

ideas that arrive in a random manner according to a Poisson process, and there is an infinite sequence of possible innovations. An idea is an ordered pair of numbers, with the first component representing the quality improvement of the innovation, and the second component the investment required to realize the innovation. If the investment embodied in the idea is not made, the idea is lost. There are no strategic interactions among firms, and a patent ceases to exist either when it reaches the end of its statutory life or when it is displaced by a better and non-infringing product.

The purported goal of a patent policy is to encourage technological progress. Yet, its impact on the pace of technological progress has largely been ignored by researchers in this field. Compared to the models of R&D in the literature on trade and technological progress, the model of R&D found in the literature of optimal patent design is much less sophisticated. In the research program on trade and technological progress of Eaton and Kortum (2001), ideas (discoveries) arrive in time according to a Poisson process whose parameter depends on the stock of knowledge – taken to be the cumulative effort of research workers. An idea is a draw from a Pareto distribution, and the draw represents the efficiency of producing the good by the technology just discovered. Using the stock of knowledge to represent the parameter of the Poisson process that characterizes the arrival in time of ideas does capture the effect of cumulative knowledge gained from R&D activities. However, this aggregate variable is not capable of capturing the impact of information externalities. To formalize the information externalities in a rigorous manner, we need to descend to a more disaggregate level and model how the findings from the R&D activities of one research organization can have a positive impact on the R&D activities of other research organizations. The objective of our paper is to fill part of this lacuna in the literature.

To fix ideas, and also to rigorously formalize the concept of information externalities, we have chosen to model R&D activities as the process of searching for a drug to treat a disease, with R&D activities being modelled from the perspective of the theory

of optimal search and where the search is being conducted on chemical compounds housed in distinct chemical libraries. The main reason for formulating the R&D process in the context of the pharmaceutical industry is because much knowledge has been gained in the process of drug discovery, and this allows for a less abstract modeling of the R&D process.

In the model we formulate, it is a given that several brands of a drug are available at any instant for treating a single disease, and that pharmaceutical firms are carrying out R&D activities to find a new drug to treat the disease. It is assumed that the patent on each brand is held by a single pharmaceutical firm, and that after the patent for a brand has expired, the brand can be manufactured as a generic drug by a competitive fringe. Furthermore, pharmaceutical firms behave strategically both in the product market and in R&D. Under this structure, three scenarios are discussed: the first scenario refers to the case where all chemical compounds in all chemical libraries are already screened. The second scenario discusses the case where the set of unscreened chemical compounds include only one element. Under this scenario it is shown that the firm which holds the earlier-expiring patent only chooses to screen the last compound, when the patent it holds expires, if the expected discounted payoff net of R&D costs yielded by this action is positive. The expected discounted payoff net of R&D costs obtained by this firm is then decreasing in the cost of screening; increasing in the cumulative quality discovered in the past screenings of the compounds in the chemical library, and decreasing in the number of past screenings carried out in this chemical library. The payoff is also higher if the marginal cost of the drug manufactured from the last compound is lower, and higher if the qualities (marginal costs) of all the other brands – generic drugs as well as the brands whose patents have not expired – are lower (higher). It is also shown that if the firm which holds the earlier-expiring patent chooses not to screen the last compound when the patent it holds expires, the firm which holds the patent on the rival brand will not choose to screen the last compound when its patent expires. The expected discounted payoff earned by the rival brand is shown to be decreasing in the cumulative quality

discovered in the chemical library, and increasing in the number of past screenings in the chemical library; increasing (decreasing) in the quality (marginal cost) of the brand and decreasing (increasing) in the qualities (marginal costs) of the generic drugs as well as the brand whose patent has not expired.

The third scenario refers to the case where the set of unscreened chemical compounds includes two elements. These elements may be located in two different chemical libraries, or both being located in the same chemical library. Under the first case the analysis suggests that if neither compound is screened by the firm which holds the earlier-expiring patent when it stands alone as the single remaining compound, then it will not be screened either by this firm when it is one of the two remaining compounds. It is also shown that the expected discounted payoff net of R&D costs that the firm which holds the earlier-expiring patent obtains by screening each of the last two compounds is positive, then, every other thing equal, the compound with the lower screening cost should be screened first, the compound from which a new drug with much higher marginal cost is manufactured should not be screened first, and the compound with the potential quality that is stochastically much larger should be screened first. When the elements are both located in the same chemical library the penultimate compound in the chemical library is screened by the firm which holds the earlier-expiring patent, and if the revealed quality of the penultimate compound, is particularly high, then the probability that the quality of the last compound in the chemical library is higher than the quality of the penultimate compound is particularly low, which means the last compound will not be screened.

Drug discovery is a long and arduous process.<sup>12</sup> A pharmaceutical company employs thousands of researchers to find new drugs that can help people with various diseases. In spite of all the progress made in science and technology, drug discovery is still a

---

<sup>12</sup> Our account of drug discovery is based upon the article “ISOA/ARF Drug Development Tutorial,” by Jens Eckstein found at <http://www.alzforum.org/drg/tut/ISOATutorial.pdf> and the article “Drug discovery,” by Citizendium found at [http://en.citizendium.org/wiki/Drug\\_discovery](http://en.citizendium.org/wiki/Drug_discovery).

lengthy and costly process. In 2000 the global R&D investment was \$US 55 billion.<sup>13</sup> On average it takes up to 15 years to develop and test a new drug, and the required investments might rise to \$US 600 million.

In the past, drugs were discovered either by identifying active ingredients in traditional remedies or by serendipity. In the new paradigm of drug discovery, researchers first try to understand how a disease is controlled at the molecular and physiological levels. The knowledge gained at this stage helps identify a target, which is the specific cellular or molecular structure involved in the pathology. Next, through a process known as high-throughput screening (HTS) large libraries of chemicals are tested for their ability to modify the target. It is estimated that only about 1 in 5000 chemicals thus tested results in a lead. Through miniaturization and robotics, it is now possible to screen millions of compounds against targets in a very short period of time. Central to the new paradigm of drug discovery is the explosion of computational techniques that help analyze enormous volumes of data, prioritize HTS hits, and guide lead optimization. If a pharmaceutical company decides to develop a lead into a drug, the drug thus developed must go through a series of clinical trials before it can be approved by the authorities. In order for a drug to be approved it must exhibit therapeutic effects against the target and prove non-toxic to the patient.

Novel chemical structures (new drugs) can be found among natural products or from libraries of synthetic compounds created by combinatorial chemistry.<sup>14</sup> The vast majority of traditionally used crude drugs in Western medicine come from roots, leaves, and flowers of plants, and a pool of information about the potential of plant species – as a source of novel chemical structure – has been accumulated. The knowledge gained from ethno-botany can prove to be valuable in narrowing down the search for novel chemical structures among plants. Microbes compete for living space

---

<sup>13</sup> See Toully *et al.* (2002).

<sup>14</sup> In the past, chemists traditionally synthesized one compound at a time. With the modern technique of combinatorial chemistry, a large number of structurally distinct molecules can be synthesized at a time, and then submitted for pharmacological assay.

and food. The chemical structures that they have developed to compete against other species – because they have been tested and proved to be effective in their struggles for survival – can be sources of new drugs. Recently, some novel chemical structures have been found among marine invertebrates.

A guiding principle in the search for a novel chemical structure is the concept of chemical diversity,<sup>15</sup> which represents the current knowledge concerning the distribution of chemical compounds – based on their physicochemical characteristics – in the chemical space. An efficient search strategy, which tells us where to search for a novel chemical structure at any point in time, whether to continue or stop the search, or whether to go to another source to search for the desired novel chemical structure should exploit our knowledge concerning the chemical diversity of the various sources. Furthermore, because R&D in the pharmaceutical industry is profit-driven, the search for a novel chemical structure also depends on economics – search costs, production costs, competition from rival firms, and the appropriability of the discoveries.

The paper is organized as follows. In Section 2, the model, which is a dynamic model in continuous time, is presented. In conventional models of patent design, only one brand – the brand with the highest quality – is available on the market at any time. Furthermore, demand is completely inelastic: each consumer, regardless of income and regardless of the price of the brand offered on the market, buys exactly one unit of the brand. In the model we formulate, several differentiated products – the products whose patents are still in force and the products whose patents have expired – are available at any time on the market. Furthermore, the demand for a brand depends on income, its own price, and the prices of the other brands. In our model,

---

<sup>15</sup> Chemical diversity is an important concept in the discipline known as chemo-informatics, which is the application of informatics methods to solve chemical problems. Chemo-informatics involves the study of chemical information or molecular similarity and the development of computational methods for the identification and optimization of active compounds to the design of smart chemical databases and comprehensive computational infrastructures for interdisciplinary pharmaceutical research.

the pharmaceutical firms with an active drug discovery program behave strategically in both R&D and in the product market. The model also contains a competitive fringe, which manufactures the brands whose patents have expired. In Section 3, some comparative static results concerning the impact of the cost and the quality of newly discovered drugs on the market are presented. The analysis of R&D is presented in Section 4. Some concluding remarks are given in Section 5.

### 3.2 The Model

Time is continuous, and is denoted by  $t, t \geq 0$ . The model deals with the case of a single disease, and in the model there are  $I$  pharmaceutical firms indexed by  $i, i = 1, \dots, I$ , where  $I$  is a positive integer, with  $I \geq 2$ . Each of these firms has an active R&D program to find a new drug for treating the disease. There is also a competitive fringe, which produces generic drugs, using the technical information disclosed in the patent of a brand after the patent on the brand has expired.

#### 3.2.1 The R&D Process

The elucidation of the structure of a chemical compound serves to avoid the rediscovery of a chemical agent that is already known for its structure and chemical activities. A chemical compound can be identified by its mass/charge ratio after ionization with the help of mass spectrometry. The technique known as nuclear magnetic resonance (NMR) can be used to obtain information about individual atoms of a compound, which allows for a detailed reconstruction of the architecture of the compound. Most software used to address the chemical diversity of a population of compounds describe each compound in the population by its molecular fingerprint, which is a bit-string consisting of 0's and 1's that represent the answers to yes-no questions about the existence or absence of sub-structural features in the molecular

structure of the compound in question.<sup>16</sup> Molecular fingerprints typically consist of hundreds, or even thousands, of bits. Thus, a 1000-bit fingerprint is a point in a 1000-dimensional chemical space. The Tanimoto dissimilarity index, which is  $1 - \text{the Tanimoto similarity index}$ , is often used to express the dissimilarity, or the distance,<sup>17</sup> between two compounds in a chemical space. In such a chemical space, similar compounds are expected to be located close to each other, while dissimilar compounds are located far apart. A much handier way to describe a compound is the BCUT<sup>18</sup> approach, which represents a compound as a point in a 6-dimensional Euclidean space. Adopting the philosophy of the BCUT approach, we shall characterize a chemical compound as a point in a low-dimensional Euclidean space, and the Euclidean distance between two compounds in such a chemical space can then be taken as how dissimilar they are as far as their chemical structures are concerned.

The search for a novel chemical structure to treat the disease in question is conducted among  $J$  sources, or libraries of chemical compounds, where  $J$  is a positive integer, with  $J \geq 2$ . For example, the search for a novel chemical structure might be conducted among plants (one source) or in the world of bacteria (another source). For each  $j = 1, \dots, J$ , let  $X_j = \{x_{j,1}, \dots, x_{j,N_j}\}$  represent the set of chemical compounds – assumed to be distinct – that are housed in the  $j$ th library, with  $N_j$  being the number of compounds this library contains. Let  $|x_{j,n} - x_{j',n'}|$  denote the Euclidean distance between two compounds  $x_{j,n}$  and  $x_{j',n'}$  in the chemical space. If the two compounds are distinct, then their distance is positive, and the more dissimilar they are the greater will be their distance.

In what follows, we shall assume that the effectiveness of a compound, say  $x_{j,n}$ , in treating the disease in question can be represented by an index, say  $q_{j,n}$ , that we call

---

<sup>16</sup> Our account of representing a compound in the chemical space is based on Pearlman (2009).

<sup>17</sup> This distance is non-Euclidean.

<sup>18</sup> Pearlman, op cit.

the quality of the chemical compound. The ordered pair  $(x_{j,n}, q_{j,n})$  will be referred to as a drug. Before a drug can be manufactured, investments must be made to construct a production plant. To concentrate on the R&D activities carried out by the firms, we choose to ignore capital investments, and simply assume that after a drug, say  $(x_{j,n}, q_{j,n})$ , has been discovered, any volume of the drug can be manufactured at a constant marginal cost, say  $c_{j,n}$ . The list  $(x_{j,n}, q_{j,n}, c_{j,n})$  will be referred to as a brand of the drug used to treat the disease in question. Note that for two brands, say  $(x_{j,n}, q_{j,n}, c_{j,n})$  and  $(x_{j',n'}, q_{j',n'}, c_{j',n'})$ , if  $x_{j,n} = x_{j',n'}$ , then it is necessary that  $q_{j,n} = q_{j',n'}$  and  $c_{j,n} = c_{j',n'}$ .

Consider a brand, say  $(x_{j,n}, q_{j,n}, c_{j,n})$ . Obviously, both  $q_{j,n}$  and  $c_{j,n}$  depend on  $x_{j,n}$ . For our purpose, we shall assume that the location of  $x_{j,n}$  in the chemical space is known and that  $c_{j,n}$ , the marginal cost of manufacturing this compound to treat the disease in question is also known. As far as the value of  $q_{j,n}$  is concerned, it is not known before the compound is screened, and thus can only be represented by a probability distribution that depends on the existing stock of knowledge.

Conceptually, we can characterize the qualities of the chemical compounds in a library, say the  $j$ th source, by a cumulative distribution function on  $(q_{j,1}, \dots, q_{j,N_j})$ , conditioned on  $(x_{j,1}, \dots, x_{j,N_j})$ . However, because of the enormous number of compounds in a chemical library, and because of the lack of knowledge on the potential of these compounds, it is extremely difficult to postulate such a distribution function for each source. Furthermore, given that a large number – about 5000 to be a little more specific – of compounds must be screened before a lead is found, we shall assume that at any point in time all the compounds in a source that have not been screened have the same potential in the following sense. First, within each source, the cost for screening a compound is the same from one compound to the next. We shall let  $\gamma_j, j = 1, \dots, J$ , denote the cost of screening a compound in the  $j$ th source. Second,

within each source, say  $j$ , the random variables  $q_{j,1}, \dots, q_{j,N_j}$  are independently and identically distributed. These assumptions imply that the order in which the compounds in a source are screened is immaterial. Hence, we can assume that the compounds in a source are screened sequentially in the increasing order. Also, to concentrate on the information externalities of the R&D process, we shall ignore its time-consuming nature, and make the simplifying assumption that the time it takes to screen a compound against a target is negligible.

We can model  $q_{j,n}, j = 1, \dots, J, n = 1, \dots, N_j$ , either as a continuous or a discrete random variable. For analytical convenience as well as for intuitive economic interpretation, we choose to model  $q_{j,n}$  as a discrete random variable. More specifically, for each  $j = 1, \dots, J$ , the potential qualities  $q_{j,1}, \dots, q_{j,N_j}$  are Poisson random variables with parameter  $\lambda_j$ , i.e., their common probability mass function is given by

$$f[q_{j,n}|\lambda_j] = \frac{\lambda_j^{q_{j,n}}}{(q_{j,n})!} e^{-\lambda_j}, \quad q_{j,n} = 0, 1, \dots, \quad (n = 1, \dots, N_j).$$

Observe that  $e^{-\lambda_j}$  represents the probability that a compound in source  $j$  has zero quality (completely ineffective in treating the disease). The Poisson process has been used by researchers, such as Arrow and Chang (1982), Chow (1981), Quyen (1991), and Cairns and Quyen (1998) to model discoveries in mineral exploration.

To model learning and the information externalities generated in the search for new drugs, we follow Quyen (1991) and adopt the approach of Bayesian statistical decision theory. Thus, for each  $j = 1, \dots, J$ , we postulate a prior density for  $\lambda_j$ , and then revise it repeatedly via the Bayes formula, using the most recent data available on screening. We assume that at the beginning  $\lambda_j$  has a gamma distribution with parameters  $\alpha_j$  and  $\beta_j$ , say

$$g_{\alpha_j, \beta_j}[\lambda_j] = \frac{1}{\Gamma(\alpha_j)\beta_j^{\alpha_j}} e^{-\lambda_j/\beta_j} \lambda_j^{\alpha_j-1}, \lambda_j > 0.$$

In the statistical literature, the gamma density is known as a conjugate distribution of the Poisson process. The choice of a conjugate prior, in addition to making the computations much simpler, allows one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, with the parameters updated by the sample information. Indeed, let

$$f_{\alpha_j, \beta_j}[q_{j,n}] = \int_0^\infty f[q_{j,n}|\lambda_j] g_{\alpha_j, \beta_j}[\lambda_j] d\lambda_j, \quad q_{j,n} = 0, 1, \dots, \quad (n = 1, \dots, N_j).$$

be the marginal density of  $q_{j,n}$ , given the parameters  $(\alpha_j, \beta_j)$  that characterize the prior density of  $\lambda_j$ . The conditional density of  $\lambda_j$ , given  $q_{j,n}$ , is

$$\begin{aligned} (1) \quad f[\lambda_j|q_{j,n}] &= \frac{g_{\alpha_j, \beta_j}[\lambda_j] f[q_{j,n}|\lambda_j]}{f_{\alpha_j, \beta_j}[q_{j,n}]} = \frac{e^{-\frac{\lambda_j}{\beta_j}} \lambda_j^{\alpha_j-1} e^{-\lambda_j} \lambda_j^{q_{j,n}}}{\Gamma[\alpha_j] \beta_j^{\alpha_j} [q_{j,n}]! f_{\alpha_j, \beta_j}[q_{j,n}]} \\ &= g_{\alpha'_j, \beta'_j}[\lambda_j] \frac{\Gamma[\alpha_j + q_{j,n}] \left(\frac{1}{1 + 1/\beta_j}\right)^{\alpha_j + q_{j,n}}}{\Gamma[\alpha_j] \beta_j^{\alpha_j} [q_{j,n}]! f_{\alpha_j, \beta_j}[q_{j,n}]}, \end{aligned}$$

where  $g_{\alpha'_j, \beta'_j}[\lambda_j]$  is the gamma density with parameters  $\alpha'_j = \alpha_j + q_{j,n}$ ,  $\beta'_j = \frac{1}{1 + 1/\beta_j}$ .

Integrating (1) with respect to  $\lambda_j$  from 0 to  $\infty$ , we obtain

$$(2) \quad \frac{\Gamma[\alpha_j + q_{j,n}] \left(\frac{1}{1 + 1/\beta_j}\right)^{\alpha_j + q_{j,n}}}{\Gamma[\alpha_j] \beta_j^{\alpha_j} [q_{j,n}]! f_{\alpha_j, \beta_j}[q_{j,n}]} = 1$$

Using (2) in (1), we see immediately that  $f(\lambda_j|q_{j,n}) = g_{\alpha'_j, \beta'_j}(\lambda_j)$ , i.e.,  $g_{\alpha'_j, \beta'_j}(\lambda_j)$  is the posterior density of  $\lambda_j$ , given  $q_{j,n}$ . Also, note that (2) can be rewritten as

$$(3) \quad f_{\alpha_j, \beta_j}[q_{j,n}] = \frac{\Gamma[\alpha_j + q_{j,n}] \left(\frac{1}{1 + 1/\beta_j}\right)^{\alpha_j + q_{j,n}}}{\Gamma[\alpha_j] \beta_j^{\alpha_j} [q_{j,n}]!}, \quad q_{j,n} = 0, 1, \dots, \quad (n = 1, \dots, N_j).$$

which is the marginal density of  $q_{j,n}$ , given the parameters  $(\alpha_j, \beta_j)$  of the prior density of  $\lambda_j$ . The distribution function associated with  $f_{\alpha_j, \beta_j}[q_{j,n}]$  is

$$(4) \quad F_{\alpha_j, \beta_j}[q] = \sum_{q'=0}^q f_{\alpha_j, \beta_j}[q'], \quad (q = 0, 1, \dots)$$

The following lemma, due to Quyen (1991), gives some useful properties of  $F_{\alpha_j, \beta_j}[q]$ .

LEMMA 1: *The distribution function  $F_{\alpha_j, \beta_j}$  is stochastically increasing in each of the parameters  $\alpha_j$  and  $\beta_j$ . More precisely, for each given value of  $q, q = 0, 1, \dots$ , we have the following results:*

- (i) *For each given  $\beta_j$ , the map  $\alpha_j \rightarrow F_{\alpha_j, \beta_j}[q]$  is strictly decreasing. Furthermore,  $\lim_{\alpha_j \rightarrow \infty} F_{\alpha_j, \beta_j}[q] = 0$ .*
- (ii) *For each given  $\alpha_j$ , the map  $\beta_j \rightarrow F_{\alpha_j, \beta_j}(q)$  is strictly decreasing. Furthermore,  $\lim_{\beta_j \rightarrow 0} F_{\alpha_j, \beta_j}[q] = 1$ .*

Before any compound in a source, say  $X_j$ , is screened, the potential quality of a typical compound in this source is believed to be Poisson with parameters  $\lambda_j$ , which in turn has a prior density that is gamma with parameters  $(\alpha_j, \beta_j)$ . The probability mass function  $f_{\alpha_j, \beta_j}[q], n = 1, \dots, n_j$ , as represented by (3), characterizes the quality of a typical compound in  $X_j$ . Now suppose that the first compound, namely  $x_{j,1}$ , is screened for its potential as a new drug to treat the disease in question, and the screening reveals that its quality is  $q_{j,1}$ . The number of compounds in  $X_j$  that remain to be screened is now  $n_j - 1$ , and the uncertain quality of a typical remaining compound is believed to be Poisson with parameter  $\lambda_j$ , which in turn is gamma with updated parameters  $\alpha'_j = \alpha_j + q_{j,1}$  and  $\beta'_j = \frac{1}{1 + \frac{1}{\beta_j}}$ . In particular, the probability mass function that characterizes the quality of a typical remaining compound is

$f_{\alpha'_j, \beta'_j}[q_{j,n}], n = 2, \dots, N_j$ . Note that the higher is the revealed quality of the compound  $x_{j,1}$  after it has been screened against the target, the more optimistic our beliefs will be concerning the potential of a typical remaining compound in source  $X_j$ . If  $n$  compounds in  $X_j$  have been screened and the qualities of these compounds have been revealed as  $q_{j,1}, \dots, q_{j,n}$ , then the potential quality of a typical remaining compound is embodied in the distribution function  $F_{\alpha'_j[n], \beta'_j[n]}[q]$ , where we have let  $\alpha'_j[1] = \alpha_j, \beta'_j[1] = \beta_j$ , and for each  $k = 1, \dots, n$ , we have let  $\alpha'_j[k+1] = \alpha'_j[k] + q_{j,k}, \beta'_j[k+1] = 1/(1 + 1/\beta'_j[k])$ . In view of Lemma 1.(i), the higher is the sum  $q_{j,1} + \dots + q_{j,n}$  of the revealed qualities, the more favourable will be the potential quality of a typical remaining compound, and this is economically quite intuitive: the more successful is the search for a new drug in a chemical library, the higher will be the potential quality of a typical remaining compound in this chemical library.

At the beginning, if we interpret  $\alpha_j$  and  $\frac{1}{\beta_j}$ , respectively, as the *cumulative quality* revealed and *the number of compounds screened* in the past, then  $\alpha'_j[n]$  and  $1/\beta'_j[n]$  represent, respectively, the cumulative revealed quality and the number of compounds that have been screened – past and present – after  $n$  more compounds have been screened. The list  $((x_{j,1}, q_{j,1}, c_{j,1}), \dots, (x_{j,n}, q_{j,n}, c_{j,n}))$  represents the history of the search in  $X_j$ , and the ordered pair  $(\alpha'_j[n], \beta'_j[n])$  embodies all the learning from the R&D activities of all the firms in the industry if the information generated in their search programs is pooled.

### 3.2.2 Patent Breadth

As far as the granting of a patent is concerned, how novel a new drug is obviously depends on its distance from the patented brand in the chemical space. Furthermore, it is also reasonable to consider a new drug as not infringing on a patented brand if it comes from a source that is different from the source from which the patented brand

is manufactured. Pig insulin, cow insulin, and human insulin produced by bacteria are all various brands of insulin used to treat diabetics. Insulin can be produced from plants<sup>19</sup> or fungi.<sup>20</sup> Recently, Symbiosis,<sup>21</sup> a Calgary-based company, claimed to have developed a technology for producing insulin from the safflower at a cost of 40% lower than that of traditional processes, and that the insulin produced from the safflower is virtually indistinguishable from human insulin. Insulin produced from the safflower thus does not infringe on the patent on human insulin produced by bacteria according to the recombinant DNA technology. Thus, for patent purposes, the distance – or novelty – between two brands, say  $(x_{j,n}, q_{j,n}, c_{j,n})$  and  $(x_{j',n'}, q_{j',n'}, c_{j',n'})$ , can be defined in the following manner

$$(5) \quad d\left((x_{j,n}, q_{j,n}, c_{j,n}), (x_{j',n'}, q_{j',n'}, c_{j',n'})\right) \\ = \epsilon_1 |x_{j,n} - x_{j',n'}| + \epsilon_2 |q_{j,n} - q_{j',n'}| + \epsilon_3 |c_{j,n} - c_{j',n'}| + \delta_{jj'},$$

where  $\epsilon_1, \epsilon_2, \epsilon_3$  are positive parameters strictly smaller than 1, and  $\delta_{jj'} = 1$  if  $j \neq j'$ ,  $\delta_{jj'} = 0$  if  $j = j'$ . Note that the distance between two brands, as defined by (5) is a weighted sum of (i) their distance in the chemical space, (ii) their quality differential, (iii) their cost differential, and (iv)  $\delta_{jj'}$ , a distance between two sources, with the distance being assigned the value 1 if they are distinct and the value 0 if they are the same. Note that the distance defined by (5) captures both product innovation – the first, second, and fourth term on the right-hand side of (5) – and process innovation – the third term on the right-hand side of (5).

Using the distance between two brands, as represented by (5), we can define the breadth – or more precisely, the leading breadth – of a patent as a number  $B, 0 < B < 1$ , such that a new brand, say  $(x_{j',n'}, q_{j',n'}, c_{j',n'})$ , is judged not to infringe

---

<sup>19</sup>“Commercial Production of Insulin and Insulin-like Proteins in Plants,” United States Patent 7393998.

<sup>20</sup> “Process for the Production of Insulin by Genetically Transformed Fungal Cells,” United States Patent 408 2613.

<sup>21</sup> <http://www.DefeatDiabetes.org>.

on a patented brand, say  $(x_{j,n}, q_{j,n}, c_{j,n})$ , if  $q_{j',n'} \geq q_{j,n}$  and  $d((x_{j,n}, q_{j,n}, c_{j,n}), (x_{j',n'}, q_{j',n'}, c_{j',n'})) > B$ . In the literature on patent design, another concept of patent breadth – lagging breadth – is also proposed. The difference between the two concepts of breadth is that the leading breadth protects the patented product from the competition by superior products, while the lagging breadth protects the patented product from the competition by inferior products. An immediate implication of the adoption of a leading breadth is that the quality of the most recently patented brand rises through time. For a more detailed analysis of these two concepts of patent breadth, the reader can consult O'Donoghue *et al.*, op cit.

We have characterized a brand as a list  $(x_{j,n}, q_{j,n}, c_{j,n})$ , which describes, respectively, its location in the chemical space, its quality in treating the disease in question, and its marginal cost (the cost of manufacturing one unit of this chemical compound). For economic considerations, it is also necessary to add another dimension to the characterization of a brand: the date on which the brand is patented. Thus, we shall from now on represent a brand by a list,  $\omega = (x_{j,n}, q_{j,n}, c_{j,n}, \tau_{j,n})$ , where  $\tau_{j,n}$  is the date on which the brand is patented. Also, we let  $L$  denote the length of a patent.

### 3.2.3 Preferences

For each  $t \geq 0$ , let  $\Omega[t]$  denote the set of brands that are available on the market at time  $t$ . The set  $\Omega[t]$  consists of the brands whose patents have expired and are now manufactured by the fringe as well as the brands whose patents are still in force. In what follows, we shall denote by  $\Omega^-[t]$  the totality of brands whose patents have expired by time  $t$ . Because a generic drug must compete against the leading brands and against the other generic drugs, not all the brands whose patents have expired will be marketed by the fringe. Thus, only those brands in  $\Omega^-[t]$  with a sufficiently low production cost and a sufficiently high quality will be marketed by the fringe. However, without any loss of generality we can assume that  $\Omega^-[t]$  is a subset of  $\Omega[t]$

by interpreting an element of  $\Omega^-[t]$  that is not marketed as a generic drug that the fringe is willing to supply, but there is no demand for it due to its high price or low quality. Note that the set  $\Omega^-[t]$  gets larger with time as more and more brands whose patents have expired are added to it.

The representative consumer consumes a numéraire good and various brands of the drug available on the market, and her preferences are represented by the following Cobb-Douglas utility function:

$$(6) \quad u(y_0, (y_\omega)_{\omega \in \Omega[t]}) = y_0^{1-\eta} \left( e^{\sum_{\omega \in \Omega[t]} \text{Log}[q_{j,n} y_\omega + 1]} \right)^\eta.$$

In (6),  $0 < \eta < 1$ , is a parameter,  $y_0$  is the consumption of the numéraire good, and  $y_\omega$  is the consumption of the brand  $\omega = (x_{j,n}, q_{j,n}, c_{j,n}, \tau_{j,n}) \in \Omega[t]$ .

The representative consumer chooses the consumption bundle  $(y_0, (y_\omega)_{\omega \in \Omega[t]})$  that maximizes (6) subject to the following budget constraint:

$$(7) \quad y_0 + \sum_{\omega \in \Omega[t]} p_\omega y_\omega = m[t].$$

In (7),  $p_\omega$  denotes the price of the brand  $\omega$ .

If we interpret the quality of a drug as its therapeutic contents, then  $\bar{y}_\omega = q_{j,n} y_\omega$  represents the effective consumption of the brand  $\omega$ . Let  $\bar{p}_\omega = p_\omega / q_{j,n}$  and  $\bar{c}_{j,n} = c_{j,n} / q_{j,n}$  denote, respectively, the price and the marginal cost – normalized by its own quality – of the brand  $\omega$ . As defined,  $\bar{p}_\omega$  and  $\bar{c}_{j,n}$  represent, respectively, the price and the marginal cost of effective consumption. In what follows, we will simply refer to  $\bar{p}_\omega$  and  $\bar{c}_{j,n}$  as the effective price and the effective marginal cost of the brand  $\omega$ , respectively. Adopting the interpretation of quality as the therapeutic contents of a drug, we can restate the problem faced by the representative consumer under the following simpler form:

$$(8) \quad \max_{(y_0, (\bar{y}_\omega)_{\omega \in \Omega[t]})} y_0^{1-\eta} \left( e^{\sum_{\omega \in \Omega[t]} \text{Log}[\bar{y}_\omega + 1]} \right)^\eta$$

subject to

$$(9) \quad y_0 + \sum_{\omega \in \Omega[t]} \bar{p}_\omega \bar{y}_\omega = m[t].$$

This is a simple utility maximization problem whose solution is given by

$$\begin{aligned} (10) \quad \bar{y}_\omega &= \frac{\eta}{1-\eta+\eta|\Omega[t]|} \frac{m[t] + \sum_{\omega' \in \Omega[t]} \bar{p}_{\omega'}}{\bar{p}_\omega} - 1 \\ &= \frac{\eta}{1-\eta+\eta|\Omega[t]|} \left( \frac{m[t] + \sum_{\omega' \in (\Omega[t]-\omega)} \bar{p}_{\omega'}}{\bar{p}_\omega} + 1 \right) - 1 \\ &= \frac{\eta}{1-\eta+\eta|\Omega[t]|} \left( \frac{m[t] + \sum_{\omega' \in (\Omega[t]-\omega)} \bar{p}_{\omega'}}{\bar{p}_\omega} \right) - \left( 1 - \frac{\eta}{1-\eta+\eta|\Omega[t]|} \right) \\ &= \frac{\eta}{1-\eta+\eta|\Omega[t]|} \left( \frac{m[t] + \sum_{\omega' \in (\Omega[t]-\omega)} \bar{p}_{\omega'}}{\bar{p}_\omega} \right) - \frac{1-2\eta+\eta|\Omega[t]|}{1-\eta+\eta|\Omega[t]|} \\ &= \frac{\kappa_1 m_\omega[t]}{\bar{p}_\omega} - \kappa_0, \quad (\omega \in \Omega[t]). \end{aligned}$$

where  $|\Omega[t]|$  denotes the number of elements of  $\Omega[t]$ , and

$$(11) \quad m_\omega[t] = m[t] + \sum_{\omega' \in (\Omega[t]-\omega)} \bar{p}_{\omega'},$$

$$(12) \quad \kappa_0 = \frac{1-2\eta+\eta|\Omega[t]|}{1-\eta+\eta|\Omega[t]|}, \kappa_1 = \frac{\eta}{1-\eta+\eta|\Omega[t]|}.$$

### 3.3 The Prices of Drugs

In practice, a pharmaceutical firm might market several brands of the drug at any time. However, to keep the exposition from becoming overburdened with notations, we shall make the simplifying assumption that if a pharmaceutical firm chooses to market a certain brand of the drug that it has patented, then it will market this brand exclusively during the statutory life of the patent on this brand. Because once the patent on a brand has expired, it will be marketed as a generic drug by the competitive fringe at cost, the firm that held the patent on the brand will not make any profit by continuing to market this brand. Thus, we shall assume that when this happens, the firm will stop marketing the brand, and must search for a new brand of

the drug if it wishes to stay in the market. For the brands whose patents have not yet expired, their prices will be strategically set by the holders of these patents. To study the strategic behaviour of these firms, we shall consider the simple case in which the set of brands whose patents have not yet expired at time  $t$  consists of only two brands, say  $\omega' = (x_{j',n'}, q_{j',n'}, c_{j',n'}, \tau_{j',n'})$  and  $\omega'' = (x_{j'',n''}, q_{j'',n''}, c_{j'',n''}, \tau_{j'',n''})$ , and that these two brands are held by two different firms. Analyzing the case of more than two pharmaceutical firms that are active in the search for a new drug will force us to consider the numerous combinations of dates on which the drugs discovered by these firms are patented. While this adds more realism to the model, it does not yield any more insights into the R&D process. Also, for simplicity we consider in this section the simple case of a single generic drug, say  $\omega = (x_{j,n}, q_{j,n}, c_{j,n}, \tau_{j,n})$ , which is manufactured and sold by the fringe at production cost, i.e.,  $\bar{p}_\omega = \bar{c}_{j,n}$ . The generalization to the case of more than one generic drug on the market can be made in a straightforward manner.

Let  $\bar{p}_{\omega'}$  ( $\bar{p}_{\omega''}$ , *resp.*) denote the effective price of the brand  $\omega'$  ( $\omega''$ , *resp.*) set at time  $t$  by the firm which holds the patent on this brand. The profit earned at time  $t$  by the brand  $\omega'$  is

$$(13) \quad (\bar{p}_{\omega'} - \bar{c}_{j',n'}) \left( \kappa_1 \frac{m_{\omega'}[t]}{\bar{p}_{\omega'}} - \kappa_0 \right).$$

Maximization of (13) with respect to  $\bar{p}_{\omega'}$ , we obtain the following reaction function for the firm that holds the patent on the brand  $\omega'$ :

$$(14) \quad \phi_{\omega'}: \bar{p}_{\omega''} \rightarrow \bar{p}_{\omega'} = \frac{\sqrt{\kappa_1} \sqrt{m[t] + \bar{p}_{\omega''} + \bar{c}_{j,n}} \sqrt{\bar{c}_{j',n'}}}{\sqrt{\kappa_0}}.$$

Note that  $\phi_{\omega'}$  is strictly increasing and strictly concave, and that its slope tends to 0 when  $\bar{p}_{\omega''} \rightarrow \infty$ .

Similarly, we obtain the following reaction function for the firm that holds the patent on the brand  $\omega''$ :

$$(15) \quad \phi_{\omega''}: \bar{p}_{\omega'} \rightarrow \bar{p}_{\omega''} = \frac{\sqrt{\kappa_1} \sqrt{m[t] + \bar{p}_{\omega'} + \bar{c}_{j,n}} \sqrt{\bar{c}_{j'',n''}}}{\sqrt{\kappa_0}}.$$

Like the reaction curve  $\phi_{\omega'}$ , the reaction curve  $\phi_{\omega''}$  is also strictly increasing, strictly concave, and its slope (with respect to the  $\bar{p}_{\omega'}$  price axis) approaches 0 when  $\bar{p}_{\omega'} \rightarrow \infty$ . The reaction curves of the two firms are depicted in Figure 6.

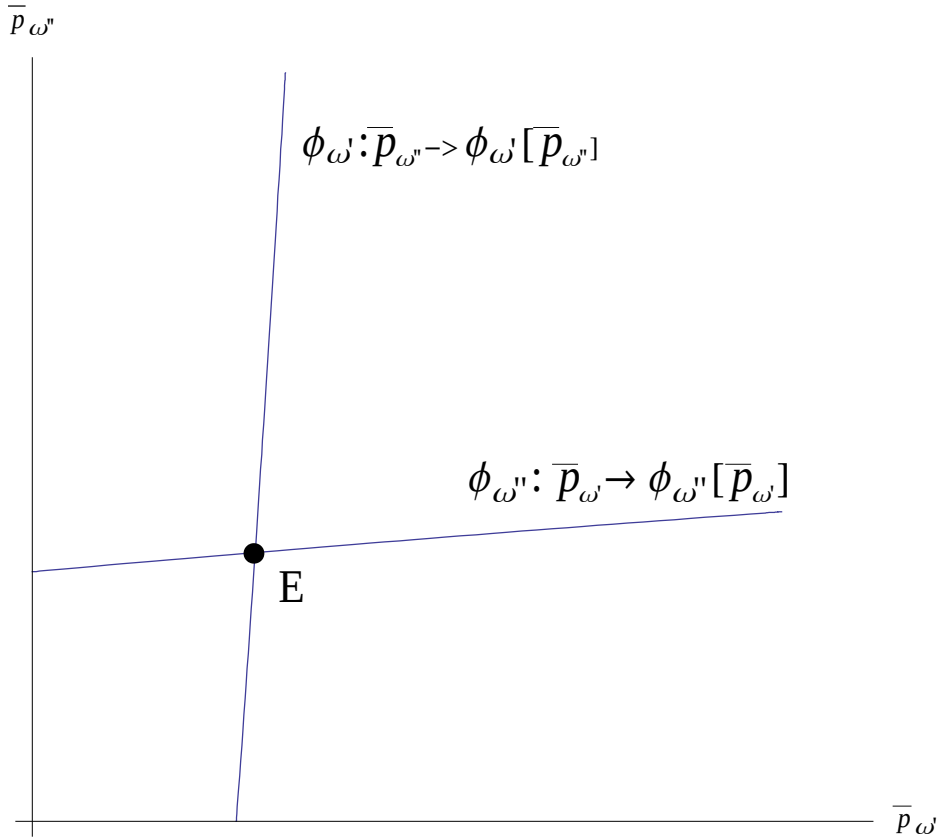


**Figure 6. The equilibrium effective prices of drugs**

The two reaction curves intersect at a unique point, which is depicted as point  $E$  in Figure 6 and which represents the effective prices set by the two firms under the equilibrium at each instant.

An alternative – and more effective – way for finding the equilibrium prices and for carrying out comparative static exercises is through the two composite maps

$$(16) \quad \Phi_{\omega'}: \bar{p}_{\omega'} \rightarrow \phi_{\omega'}[\phi_{\omega''}[\bar{p}_{\omega'}]] = \frac{1}{\sqrt{\kappa_0}} \sqrt{\kappa_1} \sqrt{\bar{c}_{j',n'}} \sqrt{m[t] + \bar{c}_{j,n} + \frac{\sqrt{\kappa_1} \sqrt{m[t] + \bar{p}_{\omega'} + \bar{c}_{j,n}} \sqrt{\bar{c}_{j'',n''}}}{\sqrt{\kappa_0}}},$$

and

(17)

$$\Phi_{\omega''}: \bar{p}_{\omega''} \rightarrow \phi_{\omega''}[\phi_{\omega'}[\bar{p}_{\omega''}]] = \frac{1}{\sqrt{\kappa_0}} \sqrt{\kappa_1} \sqrt{\bar{c}_{j'',n''}} \sqrt{m[t] + \bar{c}_{j,n} + \frac{\sqrt{\kappa_1} \sqrt{m[t] + \bar{p}_{\omega''} + \bar{c}_{j,n}} \sqrt{\bar{c}_{j',n'}}}{\sqrt{\kappa_0}}}.$$

The curve depicting the composite map  $\Phi_{\omega'}$  is depicted in Figure 7. It begins at  $\bar{p}_{\omega'} = 0$  above the 45-degree line. It is strictly increasing, strictly concave, and its slope approaches 0 when  $\bar{p}_{\omega'} \rightarrow \infty$ . Hence the curve crosses the 45-degree line at a unique point, which represents the equilibrium effective price of the brand  $\omega'$ . In the same manner, the unique fixed point of the composite map  $\Phi_{\omega''}$  represents the equilibrium effective price of the brand  $\omega''$ .

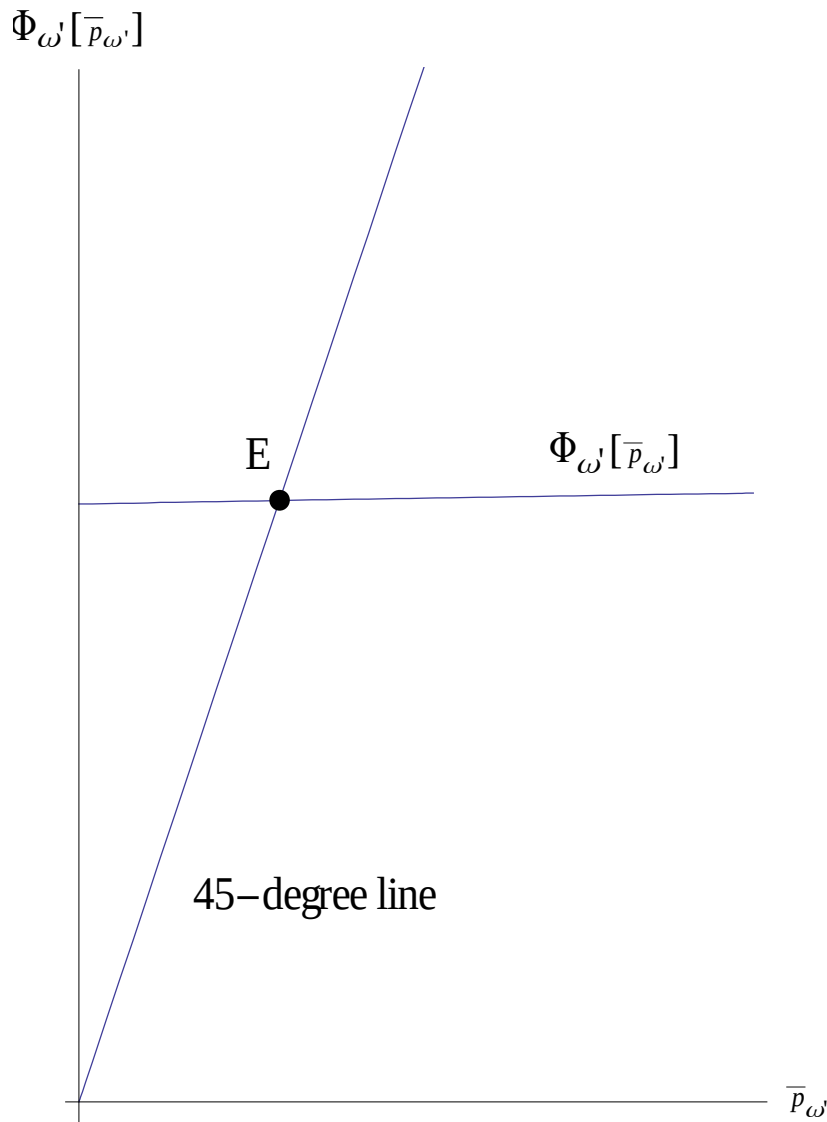


**Figure 7.** The composite map  $\Phi_{\omega'}[\bar{p}_{\omega'}] = \phi_{\omega'}[\phi_{\omega''}[\bar{p}_{\omega'}]]$

In what follows, the market share of a brand is defined as the revenue it earns relative to the total revenues earned by the industry. The following proposition gives the comparative static results of a variation in the effective marginal cost of a brand marketed by a firm that holds the patent on this brand.

PROPOSITION 1: *For a brand whose patent has not expired, the higher is its effective marginal cost,*

- (i) *the higher will be its effective price, the lower will be its market share, and the lower will be the profits it earns;*
- (ii) *the higher will be the effective price, the market share, and the profits of the rival brand;*
- (iii) *the higher will be the total revenues and market share of generic drugs.*

PROOF: The proof of Proposition 1 is given in Annex A.

The following proposition describes how the quality and the prices of generic drugs influence the pricing behaviour of the firms that hold the patents currently in force.

PROPOSITION 2: *The higher is the effective cost of the generic drugs,*

- (i) *the lower will be their revenues and their market share;*
- (ii) *the higher will be the effective price, the output, the market share, and the profits of each of the brands whose patents has not expired.*

PROOF: The proof of Proposition 2 is given in Annex B.

In what follows, we shall let  $v_{\omega'}[t, m[t], \omega', \omega'', \Omega^-[t]]$  and  $v_{\omega''}[t, m[t], \omega', \omega'', \Omega^-[t]]$  denote the profit earned at time  $t$  by the brand  $\omega'$  and the brand  $\omega''$ , respectively.

### 3.4 R&D and Information Externalities

In this section, as in Section 3.3, we shall assume that in the industry there are two firms each of which has an active drug discovery program and a competitive fringe, which markets the brands whose patents have expired as generic drugs. Also, we shall

assume that the income of the representative consumer is not varying over time; that is,  $m[t] = m = \text{constant}$ .

In a drug discovery program, a pharmaceutical firm has to decide which compound in which chemical library to screen as well as the timing of the screening. To concentrate on the information generated in the R&D activities of the pharmaceutical firms and to keep the model from becoming too complex, we shall eschew the modeling of the timing of the screenings as a strategic decision variable, and shall make two simplifying assumptions on R&D. First, a pharmaceutical firm with an active drug discovery program will only screen a chemical compound to find a new drug when the patent on the brand it currently markets expires. Second, if a pharmaceutical firm chooses not to carry out any screening when the patent on the brand it currently markets expires, it will terminate all R&D activities and exit the market. With these simplifying assumptions, the only strategic decision a pharmaceutical firm needs to consider is the chemical library and the compound in that chemical library to concentrate its efforts on. To study the evolution of technological progress in the pharmaceutical industry, we try to find the sub-game perfect Nash equilibrium of the dynamic game played by the firms in this industry. The sub-game perfect Nash equilibrium is found by backward induction on the set of chemical compounds not yet screened.

If all the compounds in all the chemical libraries have been screened, there is nothing left to discuss as far R&D is concerned, and all that needs to be done is to analyze the strategic behaviour in the product market of the holders of the patents on the brands which have not yet expired. This task was carried out in Section 3.3. So, we begin the backward induction with the case in which the set of chemical compounds that have not been screened contains only one element.

### 3.4.1 One Remaining Unscreened Chemical Compound

Suppose that among all the chemical libraries there remains only a compound that has not been screened, say the compound  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$ , which is the last compound in the  $j$ th chemical library. Also, suppose that the information obtained from the past screenings in this chemical library leads us to believe that the potential quality of this chemical compound is represented by the probability mass function  $f(\alpha_{j[N_j]}, \beta_{j[N_j]})(q_{j,N_j})$ .

Let  $\omega' = (x_{j',n'}, q_{j',n'}, c_{j',n'}, \tau_{j',n'})$  and  $\omega'' = (x_{j'',n''}, q_{j'',n''}, c_{j'',n''}, \tau_{j'',n''})$  be the two brands currently marketed by the two pharmaceutical firms that hold the patents on these brands. Without any loss of generality, we assume that  $\tau'' < \tau'$ . That is, the patent on the brand  $\omega''$  was granted before that on the brand  $\omega'$ , and thus the patent on the former brand will expire before the patent on the latter brand. According to the assumption that a pharmaceutical firm will only carry out the screenings needed to find a new drug when the patent it holds on the brand it currently markets expires, the last chemical compound will be screened by the firm which holds the patent on the brand  $\omega''$  if this firm finds the screening to be profitable.

If the pharmaceutical firm which holds the patent on the brand  $\omega''$  chooses to screen the last compound, then the newly discovered drug can only be patented if its quality is at least equal to the quality of the brand that is most recently patented, namely the brand  $\omega'$ , and if its distance from  $\omega'$  is greater than the breadth that protects the brand  $\omega'$ ; that is, if  $q_{j,N_j}$  belongs to the set

$$(18) Q_{j,N_j}[\omega'] = \left\{ q_{j,N_j} \mid q_{j,N_j} \geq q_{j',n'}, d\left((x_{j',n'}, q_{j',n'}, c_{j',n'}), (x_{j,N_j}, q_{j,N_j}, c_{j,N_j})\right) > B \right\}.$$

In this case, the discounted value – discounted to the time the patent on the brand  $\omega''$  expires – of the stream of profits earned by the newly discovered drug is given by

$$(19) \quad A_1[q_{j,N_j}] = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega^-}[s, m, \omega^-, \omega', \Omega^-[s]] ds \\ + \int_{\tau_{j',n'}+L}^{\tau_{j'',n''}+2L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega^-}[s, m, \omega^-, \Omega^-[s]] ds.$$

In (19),  $r$  is the market rate of interest, and  $\omega^- = (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}, \tau_{j,N_j})$  represents the newly discovered drug, with  $\tau_{j,N_j} = \tau_{j'',n''} + L$  as the date on which the last compound is screened, which is also the date on which the newly discovered drug is patented. Also,  $\Omega^-[s]$ , we recall, represents the set of brands whose patents have expired by time  $s$ .

The first integral in (19) represents the discounted value of the stream of profits earned by the newly discovered drug from the time of its discovery until the time the patent on the brand  $\omega'$  expires. During the time interval  $[\tau_{j'',n''} + L, \tau_{j',n'} + L)$ , the newly discovered drug has to compete against the brand  $\omega'$  marketed by the rival pharmaceutical firm which holds the patent on this brand as well as against all the generic drugs – which include its own predecessor (the brand  $\omega''$ ) and all the brands whose patents expired by time  $\tau_{j'',n''} + L$ . Under the integral sign, the expression  $v_{\omega^-}[s, m, \omega^-, \omega', \Omega^-[s]]$  represents the profit earned by the newly discovered drug at each time  $s$  during the time interval  $[\tau_{j'',n''} + L, \tau_{j',n'} + L)$ . According to Proposition 1,  $v_{\omega^-}[s, m, \omega^-, \omega', \Omega^-[s]]$  is strictly increasing (decreasing) in the marginal cost (quality) of the rival brand  $\omega'$ . According to Proposition 2,  $v_{\omega^-}[s, m, \omega^-, \omega', \Omega^-[s]]$  is strictly decreasing (increasing) in the quality (marginal cost) of each generic drug. The second integral in (19) represents the discounted value of the stream of profits earned by the newly discovered drug from time  $\tau_{j',n'} + L$  until the end of its statutory life. During this time interval, the newly discovered drug only faces competition from the generic drugs, which include the brand  $\omega'$  whose patent expired at time  $\tau_{j',n'} + L$ . The expression  $v_{\omega^-}[s, m, \omega^-, \Omega^-[s]]$  under the integral sign represents the profit

earned by the newly discovered drug at each time  $s$  during the time interval  $[\tau_{j',n'} + L, \tau_{j'',n''} + 2L)$ . Note that  $v_{\omega^-}[s, m, \omega^-, \Omega^-[s]]$  is strictly decreasing (increasing) in the quality (marginal cost) of each generic drug.

According to Proposition 1,  $A_1[q_{j,N_j}]$  is strictly increasing in  $q_{j,N_j}$ . It is also higher if the quality (marginal cost) of the brand  $\omega'$  is lower (higher). According to Proposition 2,  $A_1[q_{j,N_j}]$  is higher if qualities (marginal costs) of the generic drugs are lower (higher).

Under the event that the screening of the compound  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$  does not yield any patentable drug, the firm that holds the patent on the brand  $\omega''$  incurred the R&D cost  $\gamma_j$  without obtaining any benefits.

The discounted value of the stream of profits earned by the firm which holds the patent on the brand  $\omega''$  if it chooses to screen the last compound when this patent expires – as a function of the quality of this compound – is then given by

$$(20) \quad A_2[q_{j,N_j}] = \begin{cases} A_1[q_{j,N_j}] & \text{if } q_{j,N_j} \in Q_{j,N_j}[\omega'], \\ 0, & \text{otherwise.} \end{cases}$$

From the discussion on the properties of  $A_1[q_{j,N_j}]$ , we can assert that  $A_2[q_{j,N_j}]$  is increasing (decreasing) in  $q_{j,N_j}$  ( $c_{j,n}$ ). It is decreasing (increasing) in the qualities (marginal costs) of all the other brands – those whose patents have expired and those whose patents have not expired. The expected discounted profit net of R&D costs – with the discounting being carried back to the time the patent on the brand  $\omega''$  expires – that the firm which holds this patent obtains from by screening the last compound is then given by

$$(21) \quad A_3 = -\gamma_j + \sum_{q_{j,N_j} \geq 0} A_2[q_{j,N_j}] f(\alpha_j[N_j], \beta_j[N_j]) [q_{j,N_j}].$$

Let

(22)

$$V_{\omega''} \left[ \omega', \omega'', \Omega^- [\tau_{j'',n''} + L] \left[ \left( x_{j,N_j}, q_{j,N_j}, c_{j,N_j} \right), f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}] \right] \right] = \max\{0, A_3\}.$$

denote the expected discounted payoff net of R&D cost – discounted to the time the patent on the brand  $\omega''$  expires – that the firm which holds the patent on the brand  $\omega''$  obtains, given that (i) the patent on the brand  $\omega''$  was granted before the patent on the rival brand  $\omega'$ , and (ii) the probability mass function which characterizes the quality of the last compound is  $f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}]$ .

**PROPOSITION 3:** *The firm which holds the patent on the brand  $\omega''$  only chooses to screen the last compound when the patent it holds on the brand  $\omega''$  expires if  $A_3 > 0$ , i.e., if the expected discounted payoff net of R&D costs yielded by this action is positive. Furthermore, if this is the case, then the expected discounted payoff net of R&D costs obtained by this firm is*

- (i) *decreasing in  $\gamma_j$ , the cost of screening;*
- (ii) *increasing in  $\alpha_j[N_j]$ , the cumulative quality discovered in the past screenings of the compounds in the  $j$ th chemical library, and decreasing in  $\frac{1}{\beta_j[N_j]}$ , the number of past screenings carried out in this chemical library;*
- (iii) *decreasing in the marginal cost of the drug manufactured from the last compound.*
- (iv) *decreasing (increasing) in the qualities (marginal costs) of all the other brands – generic drugs as well as the brands whose patents have not expired.*

**PROOF:** The proof of Proposition 3 is given in Annex C.

LEMMA 2: *If the firm which holds the patent on the brand  $\omega''$  chooses not to screen the last compound when the patent it holds on the brand  $\omega''$  expires, then neither will the firm which holds the patent on the rival brand  $\omega'$  choose to screen the last compound when this patent expires.*

PROOF: The proof of Lemma 2 is given in Annex D.

Now let

$$(23) \quad V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \left[ \left( x_{j,N_j}, q_{j,N_j}, c_{j,N_j} \right), f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}] \right] \right]$$

denote the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – that the firm which holds the patent on the brand  $\omega'$  obtains, given that (i) the patent on the brand  $\omega''$  was granted before the patent on the rival brand  $\omega'$ , and (ii) the probability mass function which characterizes the quality of the last compound is  $f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}]$ . This expected discounted payoff can be computed as follows.

If the firm which holds the patent on the brand  $\omega''$  chooses to screen the last compound and obtains a patentable drug  $\omega^-$  from the screening, then the discounted value – discounted to the time the patent on the brand  $\omega''$  expires – of the stream of profits earned by the brand  $\omega'$  from time  $\tau_{j'',n''} + L$  until the end of its statutory life is

$$(24) \quad A_4[q_{j,N_j}] = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega'}[s, m, \omega^-, \omega', \Omega^-[s]] ds.$$

On the other hand, under the event that the firm which holds the patent on the brand  $\omega''$  fails to obtain a patentable drug from screening the last compound, the discounted value – discounted to the time the patent on the brand  $\omega''$  expires – of the stream of profits earned by the brand  $\omega'$  from time  $\tau_{j'',n''} + L$  until the end of its statutory life is

$$(25) \quad A_5 = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega'}[s, m, \omega', \Omega^-[s]] ds.$$

Thus, if the firm which holds the patent on the brand  $\omega''$  chooses to screen the last chemical compound when the patent on the brand  $\omega''$  expires, then the expected discounted value of the stream of profits earned by the brand  $\omega'$  from time  $\tau_{j'',n''} + L$  until the end of its statutory life – with the discounting being carried back to the time the patent on the brand  $\omega''$  expires – is given by

$$(26) \quad A_6 = \sum_{q_{j,N_j} \in Q_{j,N_j}[\omega']} A_4[q_{j,N_j}] f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}] \\ + A_5 \left( 1 - \sum_{q_{j,N_j} \in Q_{j,N_j}[\omega']} f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}] \right).$$

If the firm which holds the patent on the brand  $\omega''$  chooses not to screen the last chemical compound, then according to Lemma 2, the last compound will not be screened by the firm which holds the patent on the brand  $\omega'$  either, when the patent on the brand  $\omega'$  expires. Under this scenario, the expected discounted value – discounted to the time the patent on the brand  $\omega''$  expires – of the stream of profits earned by the brand  $\omega'$  is also given by  $A_5$ .

We are now ready to give the explicit expression for (23) as follows:

$$(27) \quad V_{\omega'}[\omega', \omega'', \Omega^-[ \tau_{j'',n''} + L ] \left[ (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}] \right] = \\ \begin{cases} A_6 & \text{if the firm which holds the patent on the brand } \omega'' \text{ screens the last compound,} \\ A_5 & \text{if the firm which holds the patent on the brand } \omega'' \text{ does not screen the last compound.} \end{cases}$$

**PROPOSITION 4:** *The expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – that is earned by the firm which holds the patent on the brand  $\omega'$  from time  $\tau_{j'',n''} + L$  until the end of the statutory life of the patent on the brand  $\omega'$  is*

- (i) *decreasing in  $\alpha_j[N_j]$ , the cumulative quality discovered in the  $j$ th chemical library, and increasing in  $\frac{1}{\beta_j[N_j]}$ , the number of past screenings in the  $j$ th chemical library;*
- (ii) *increasing (decreasing) in the quality (marginal cost) of the brand  $\omega'$*
- (iii) *decreasing (increasing) in the qualities (marginal costs) of the generic drugs as well as the brands whose patents have not expired.*

PROOF: The proof of Proposition 4 is given in Annex E.

### 3.4.2 Two Remaining Unscreened Chemical Compounds

Let  $\omega' = (x_{j',n'}, q_{j',n'}, c_{j',n'}, \tau_{j',n'})$  and  $\omega'' = (x_{j'',n''}, q_{j'',n''}, c_{j'',n''}, \tau_{j'',n''})$ , with  $\tau'' < \tau'$ , be the two brands currently marketed by the two pharmaceutical firms which hold the patents on these brands. Suppose that there remain only two chemical compounds that have not been screened, say  $(x_{j[\ell],n[\ell]}, q_{j[\ell],n[\ell]}, c_{j[\ell],n[\ell]}), \ell = 1, 2$ . There are two possibilities to consider: (i) the two compounds are located in two different chemical libraries, and (ii) the two compounds are located in the same chemical library. Under the first possibility, screening one compound yields only information about this compound, but no information about the other compound. Under the second possibility, screening the penultimate compound yields information about this compound and information about the last compound.

#### 3.4.2.1 The Two Chemical Compounds are Located in Two Different Chemical Libraries

Suppose that the last two compounds that have not been screened are located in two different chemical libraries. Under this scenario, each compound is the last one in a chemical library, i.e.,

$$(x_{j[\ell],n[\ell]}, q_{j[\ell],n[\ell]}, c_{j[\ell],n[\ell]}) = (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}), \quad (\ell = 1, 2),$$

with  $j[1] \neq j[2]$ . Also, suppose that the qualities of these compounds are characterized by the probability mass functions  $f_{(\alpha_{j[\ell]}, N_{j[\ell]}), (\beta_{j[\ell]}, N_{j[\ell]})} [q_{j[\ell]}, N_{j[\ell]}]$ ,  $\ell = 1, 2$ .

For each  $\ell = 1, 2$ , let

$$(28) A_7^\ell =$$

$$V_{\omega''} \left[ \omega', \omega'', \Omega^- [\tau_{j'', n''} + L] \mid (x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}}), f_{(\alpha_{j[\ell]}, N_{j[\ell]}), (\beta_{j[\ell]}, N_{j[\ell]})} [q_{j[\ell]}, N_{j[\ell]}] \right]$$

denote the expected discounted payoff net of R&D costs that the firm which holds the patent on the brand  $\omega''$  obtains under the assumption that  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  stands alone as the single remaining compound, not as one of the two remaining compounds.

LEMMA 3:

- (i) *If  $A_7^\ell \leq 0$ ,  $\ell = 1, 2$ , i.e., if neither compound is screened by the firm which holds the patent on the brand  $\omega''$  when it stands alone as the single remaining compound, then it will not be screened, either, by this firm when it is one of the two remaining compounds. The same result holds for the firm which holds the patent on the brand  $\omega'$  when this patent expires.*
- (ii) *If  $A_7^\ell > 0$ ,  $A_7^{\ell'} \leq 0$ , then the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  will be screened by the firm which holds the patent on the brand  $\omega''$  when this patent expires. Furthermore, regardless of the outcome of the screening, the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  will not be screened by the firm which holds the patent on the brand  $\omega'$  when this patent expires.*

PROOF: The proof of Lemma 3 is given in Annex F.

If  $A_7^\ell > 0$ ,  $\ell = 1, 2$ , then neither compound can be summarily rejected, and the question concerning the order in which the compounds are screened arises.

Suppose that the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  immediately after the patent on the brand  $\omega''$  has expired. Under the event that the screening leads to a patentable drug, i.e., if  $q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']$ , where

$$(29) \quad Q_{j[\ell], N_{j[\ell]}}[\omega'] = \left\{ q_{j[\ell], N_{j[\ell]}} \mid q_{j[\ell], N_{j[\ell]}} \geq q_{j', n'}, d((x_{j', n'}, q_{j', n'}, c_{j', n'}), (x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})) > B \right\},$$

then the discounted value – discounted to time the patent on the brand  $\omega''$  expires – of the stream of profits earned by the newly discovered drug during the time interval  $[\tau_{j'', n''} + L, \tau_{j', n'} + L)$  is given by

$$(30) \quad A_8^\ell [q_{j[\ell], N_{j[\ell]}}] = \int_{\tau_{j'', n''} + L}^{\tau_{j', n'} + L} e^{-r(s - \tau_{j'', n''} - L)} v_{\omega^\ell} [s, m, \omega^\ell, \omega', \Omega^-[s]] ds,$$

where  $\omega^\ell = (x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}}, \tau_{j[\ell], N_{j[\ell]}})$ , with  $\tau_{j[\ell], N_{j[\ell]}} = \tau_{j'', n''} + L$ , denotes the newly discovered drug. Furthermore, at time  $\tau_{j', n'} + L$ , when the patent on the brand  $\omega'$  expires, the two pharmaceutical firms find themselves exactly in the situation analyzed in Sub-section 3.4.1, but with their roles reversed. Now there is only one compound left for screening, namely the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$ ,  $\ell' \neq \ell$ , and the two brands marketed by the two pharmaceutical firms are  $\omega^\ell$  and  $\omega'$ , with the patent on the latter brand being granted before the patent on the former brand. Under such a scenario, the expected discounted payoff – discounted to the time the patent on the brand  $\omega'$  expires – earned by the firm which holds the patent on the brand  $\omega^\ell$  after the patent on the brand  $\omega'$  has expired is given by

$$(31) \quad A_9^\ell[q_{j[\ell],N_{j[\ell]}}] = V_{\omega^\ell} \left[ \omega^\ell, \omega', \Omega^-[\tau_{j',n'} + L] \mid (x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}}), f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}])(q_{j[\ell'],N_{j[\ell']}}) \right].$$

The discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – earned by the firm which holds the patent on the brand  $\omega''$  under the event that the screening of the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  results in a patentable drug is then given by

$$(32) \quad A_{10}^\ell[q_{j[\ell],N_{j[\ell]}}] = A_8^\ell[q_{j[\ell],N_{j[\ell]}}] + e^{-r(\tau_{j',n'} - \tau_{j'',n''})} A_9^\ell[q_{j[\ell],N_{j[\ell]}}].$$

On the other hand, if the screening of the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  is not fruitful, then the two firms also find themselves in the same situation analyzed in Sub-section 3.4.1, with  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$  being the last remaining chemical compound. The expected discounted payoff – discounted to the instant the patent on the brand  $\omega''$  expires – that this firm earns after the first unsuccessful attempt is  $A_7^{\ell'}$ , according to (28).

If the first screening was not fruitful, and if  $A_7^{\ell'} > 0$ , then there will be a surge in R&D activities: the firm which holds the patent on the brand  $\omega''$  will continue the search by screening immediately the remaining compound, namely the compound  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$ . Under this scenario,  $A_7^{\ell'}$  represents the discounted payoff net of R&D cost – discounted to the time the patent on the brand  $\omega''$  expires – that is yielded by the second screening.

The expected discounted payoff net of R&D costs – discounted to the time the patent on the brand  $\omega''$  expires – that the firm which holds the patent on the brand  $\omega''$  obtains by screening the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  is given by

(33)

$$A_{11}^{\ell} = -\gamma_{j[\ell]} + \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} A_{10}^{\ell} [q_{j[\ell], N_{j[\ell]}}] f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}}] \\ + A_7^{\ell'} \left( 1 - \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}}] \right).$$

Obviously, if  $A_{11}^{\ell} \leq 0$ , then the  $\ell$ th remaining compound will not be screened by the firm which holds the patent on the brand  $\omega''$  when this patent expires. In particular, if  $A_{11}^{\ell} \leq 0$ , for each  $\ell = 1, 2$ , then the firm which holds the patent on the brand  $\omega''$  will shut down its R&D activities and exit the market after the patent on the brand  $\omega''$  has expired.

When both  $A_{11}^{\ell}$  and  $A_{11}^{\ell'}$  are positive, it is necessary to compare them in order to find out which compound should be screened first. Because of the numerous parameters involved – the dates on which the patents on the brands  $\omega'$  and  $\omega''$  were granted; the screening costs of the two remaining unscreened compounds; the potential qualities of these compounds; and the marginal costs of the drugs manufactured from these compounds – it is difficult to determine unambiguously the sign of the payoff differential  $A_{11}^{\ell} - A_{11}^{\ell'}$ . Intuitively, we expect that the order of screening should favour the compound with lower screening cost and higher potential quality. It should also favour the compound from which a drug with a lower marginal cost could be discovered. The following proposition confirms our intuition of the influence of screening costs on the order of screening.

**PROPOSITION 5:** *Suppose that  $A_{11}^{\ell} > 0, \ell = 1, 2$ . Then, every other thing equal, the compound with the lower screening cost should be screened first.*

**PROOF:** The proof of Proposition 5 is given in Annex G.

PROPOSITION 6: Suppose that  $A_{11}^\ell > 0, \ell = 1, 2$ . Every other thing equal, the compound from which a new drug with much higher marginal cost is manufactured should not be screened first.

PROOF: The proof of Proposition 6 is given in Annex H.

PROPOSITION 7: Suppose that  $A_{11}^\ell > 0, \ell = 1, 2$ . Then every other thing equal, the compound with the potential quality that is stochastically much larger should be screened first.

PROOF: The proof of Proposition 7 is given in Annex I.

Let

(34)

$$V_{\omega''} \left[ \omega', \omega'', \Omega^-[\tau_{j'', n''} + L] \left| \left( (x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}}), f(\alpha_{j[\ell], N_{j[\ell]}}, \beta_{j[\ell], N_{j[\ell]}})[q_{j[\ell], N_{j[\ell]}}] \right)_{\ell=1}^2 \right. \right] \\ = \max \left\{ 0, (A_{11}^\ell)_{\ell=1}^2 \right\}$$

denote the expected discounted payoff for the firm which holds the patent on the brand  $\omega''$ , given that (i) the patent on the brand  $\omega''$  was granted before the patent on the brand  $\omega'$ , (ii) there are two remaining unscreened compounds,  $\left( (x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}}) \right)_{\ell=1}^2$ , which are located in two different chemical libraries, and (iii) for each  $\ell = 1, 2$ , the potential quality of the remaining unscreened compound  $\ell$  is represented by the probability mass function  $f(\alpha_{j[\ell], N_{j[\ell]}}, \beta_{j[\ell], N_{j[\ell]}})[q_{j[\ell], N_{j[\ell]}}]$ .

Let

(35)

$$V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \left| \left( (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}), f_{(\alpha_{j[\ell],N_{j[\ell]}}, \beta_{j[\ell],N_{j[\ell]}})}[q_{j[\ell],N_{j[\ell]}}] \right)_{\ell=1}^2 \right. \right]$$

denote the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$ , given that (i) the patent on the brand  $\omega''$  was granted before the patent on the brand  $\omega'$ , (ii) there are two remaining unscreened compounds,  $\left( (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}) \right)_{\ell=1}^2$ , which are located in two different chemical libraries, and (iii) for each  $\ell = 1, 2$ , the potential quality of the remaining unscreened compound  $\ell$  is represented by the probability mass function  $f_{(\alpha_{j[\ell],N_{j[\ell]}}, \beta_{j[\ell],N_{j[\ell]}})}[q_{j[\ell],N_{j[\ell]}}]$ . This expected discounted payoff can be computed as follows.

First, if the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $\left( x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}} \right)$  immediately after the patent on the brand  $\omega''$  has expired, and if the screening results in a patentable drug  $\omega^\ell = (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}, \tau_{j[\ell],N_{j[\ell]}})$ , with  $\tau_{j[\ell],N_{j[\ell]}} = \tau_{j'',n''} + L$ , then the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$  is given by

$$(36) \quad A_{12}^\ell[q_{j[\ell],N_{j[\ell]}}] = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega'}[s, m, \omega^\ell, \omega', \Omega^-[s]] ds + e^{-r(\tau_{j',n'}-\tau_{j'',n''})} \times \\ V_{\omega'} \left[ \omega^\ell, \omega', \Omega^-[\tau_{j',n'} + L] \left| (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}), f_{(\alpha_{j[\ell],N_{j[\ell]}}, \beta_{j[\ell],N_{j[\ell]}})}[q_{j[\ell],N_{j[\ell]}}] \right. \right].$$

In (44),  $\Omega^-[\tau_{j',n'} + L] = \Omega^-[\tau_{j'',n''} + L] \cup \{\omega'\}$ .

If the screening of the compound  $\left( x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}} \right)$  is not fruitful, then immediately after this screening, we have the situation analyzed in Sub-section 3.4.1,

and the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$  is given by

$$(37) \quad A_{13}^{\ell'} = V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \mid (x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}}), f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}])(q_{j[\ell'],N_{j[\ell']}}) \right].$$

Thus, under the scenario that the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  immediately after the patent on the brand  $\omega''$  has expired, then the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$  is given by

$$(38) \quad A_{14}^{\ell} = \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} A_{12}^{\ell}[q_{j[\ell],N_{j[\ell]}}] f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}])(q_{j[\ell],N_{j[\ell]}}) \\ + \left( (1 - \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}])(q_{j[\ell],N_{j[\ell]}})) \right) A_{13}^{\ell'}.$$

Second, if the firm which holds the patent on the brand  $\omega''$  chooses not to carry out any screening when the patent it holds on the brand  $\omega''$  expires, then it will shut down its R&D activities and exit the market. Under this scenario, the discounted profit – discounted to the time the patent on the brand  $\omega''$  expires – earned by the brand  $\omega'$  from time  $\tau_{j'',n''} + L$  until time  $\tau_{j',n'} + L$  is given by

$$(39) \quad A_{15} = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega'}[s, m, \omega', \Omega^-[s]] ds.$$

At time  $\tau_{j',n'} + L$ , when the patent it holds on the brand  $\omega'$  expires, this firm might choose to screen one of the remaining compounds, say the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$ . If the screening is fruitful and results in a patentable drug, say  $\omega^{\ell} = (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$ , then the expected discounted payoff – discounted to time  $\tau_{j',n'} + L$  – obtained by this firm is given by

$$(40) \quad A_{16}^{\ell}[q_{j[\ell],N_{j[\ell]}}] = \int_{\tau_{j',n'}+L}^{\tau_{j',n'}+2L} e^{-r(s-\tau_{j',n'}-L)} v_{\omega^{\ell}}[s, m, \omega^{\ell}, \Omega^{-}[s]] ds + e^{-rL} \times \\ V_{\omega^{\ell}} \left[ \Omega^{-}[\tau_{j',n'} + 2L] \left| (x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}}), f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}] ) [q_{j[\ell'],N_{j[\ell']}}] \right. \right].$$

On the other hand, if the screening of the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  is not fruitful, then the expected discounted payoff – discounted to time  $\tau_{j',n'} + L$  – obtained by the firm after the patent on the brand  $\omega'$  has expired is given by

$$(41) \quad A_{17}^{\ell'} = V_{\omega'} \left[ \Omega^{-}[\tau_{j',n'} + L] \left| (x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}}), f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}] ) [q_{j[\ell'],N_{j[\ell']}}] \right. \right].$$

The expected discounted payoff – discounted to the time the patent on the brand  $\omega'$  expires – earned by the firm which holds the patent on the brand  $\omega'$  after this patent has expired, given that (i) the firm which holds the patent on the brand  $\omega''$  chooses not to carry out any screening when this patent expires, and (ii) the firm which holds the patent on the brand  $\omega'$  chooses to screen the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  when the patent on the brand  $\omega'$  expires, is then given by

$$(42) \quad A_{18}^{\ell} = -\gamma_{j[\ell]} + \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} A_{16}^{\ell}[q_{j[\ell],N_{j[\ell]}}] f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}] \\ + \left( 1 - \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}] \right) A_{17}^{\ell'}.$$

The expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – earned by the firm which holds the patent on the brand  $\omega'$  after this patent has expired, given that (i) the firm which holds the patent on the brand  $\omega''$  chooses not to carry out any screening when this patent expires, and (ii) the firm which holds the patent on the brand  $\omega'$  chooses to screen the compound

$(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  when the patent on the brand  $\omega'$  expires, is then given by

$$(43) \quad A_{19} = A_{15} + e^{-r(\tau_{j',n'} - \tau_{j'',n''})} \max \{0, A_{18}^\ell, A_{18}^{\ell'}\}.$$

We are now ready to give the explicit expression for (35) as

(44)

$$V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \left| \left( (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}), f_{(\alpha_{j[\ell]}, \beta_{j[\ell]})} [q_{j[\ell],N_{j[\ell]}}] \right)_{\ell=1}^2 \right. \right] \\ = \begin{cases} A_{14}^\ell & \text{if the firm which holds the patent on the brand } \omega'' \text{ chooses to screen} \\ \text{the compound } (x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}}) & \text{when the patent on the brand } \omega'' \text{ expires,} \\ A_{19} & \text{if the firm which holds the patent on the brand } \omega'' \text{ chooses not to screen one of} \\ & \text{the two remaining compounds when the patent on the brand } \omega'' \text{ expires.} \end{cases}$$

### 3.4.2.2 The Two Remaining Compounds are Located in the Same Chemical Library

Suppose that the last two chemical compounds that have not been screened are both located in the  $j$ th chemical library, i.e., the two remaining compounds are  $(x_{j,N_{j-1}}, q_{j,N_{j-1}}, c_{j,N_{j-1}})$  and  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$ . Furthermore, their uncertain qualities are characterized by the same probability mass function  $f_{(\alpha_j, \beta_j)} [q_{j,N_{j-1}}]$ . The firm which holds the patent on the brand  $\omega''$  has only two possible choices when this patent expires: to screen or not to screen.

Suppose that the firm which holds the patent on the brand  $\omega''$  chooses to screen the penultimate compound immediately when the patent on the brand  $\omega''$  expires, and that the screening results in a patentable drug, say  $\omega^1 = (x_{j,N_{j-1}}, q_{j,N_{j-1}}, c_{j,N_{j-1}}, \tau_{j,N_{j-1}})$ , with  $\tau_{j,N_{j-1}} = \tau_{j'',n''} + L$ ,  $q_{j,N_{j-1}} \in Q_{j,N_{j-1}}[\omega']$ , where

$$(45) \quad Q_{j,N_j-1}[\omega'] = \left\{ q_{j,N_j-1} \mid q_{j,N_j-1} \geq q_{j',n'}, d((x_{j',n'}, q_{j',n'}, c_{j',n'}), (x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})) > B \right\}$$

then the discounted value – discounted to time the patent on the brand  $\omega''$  expires – of the stream of profits earned by the newly discovered drug during the time interval  $[\tau_{j'',n''} + L, \tau_{j',n'} + L)$  is

$$(46) \quad \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega^1}[s, m, \omega^1, \omega', \Omega^-[s]] ds,$$

Furthermore, at time  $\tau_{j',n'} + L$ , when the patent on the brand  $\omega'$  expires, the two pharmaceutical firms find themselves exactly in the situation analyzed in Sub-section 3.4.1, but with their roles reversed. Now there is only one compound left for screening, namely the compound  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$ , and the two brands marketed by the two pharmaceutical firms are  $\omega'$  and  $\omega^1$ , with the patent on the former brand being granted before that on the latter. Furthermore, the probability mass function that characterizes the uncertain quality of the compound  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$  is now revised – in light of the revealed quality of the penultimate compound – to be  $f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}]$ , where  $\alpha_j[N_j] = \alpha_j[N_j - 1] + q_{j,N_j-1}$  and  $\frac{1}{\beta_j[N_j]} = 1 + \frac{1}{\beta_j[N_j-1]}$ . Also, according to Lemma 1, the revised distribution that characterizes our beliefs concerning the uncertain quality of the last remaining compound is stochastically increasing in the quality of the newly discovered drug. Under such a scenario, the expected discounted payoff – discounted to the time the patent on the brand  $\omega'$  expires – earned by the firm which holds the patent on the brand  $\omega^1$  after the patent on the brand  $\omega'$  has expired is given by

$$(47) \quad A_{20}[q_{j,N_j-1}] = V_{\omega^1}[\omega^1, \omega', \Omega^-[ \tau_{j',n'} + L ] \mid (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}]].$$

The discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – earned by the firm which holds the patent on the brand  $\omega''$  under the event that the screening of the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  results in a patentable drug is then given by

$$(48) \quad A_{21}[q_{j,N_j-1}] = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega^1}[s, m, \omega^1, \omega', \Omega^-[s]] ds \\ + e^{-r(\tau_{j',n'}-\tau_{j'',n''})} A_{20}[q_{j,N_j-1}].$$

On the other hand, if the screening of the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  is not fruitful, then the two firms also find themselves in the same situation analyzed in Sub-section 3.4.1, with  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$  being the last chemical compound. Under this event, this firm might give up, or it might screen the last compound immediately after the first unsuccessful attempt. The expected discounted payoff – discounted to the instant the patent on the brand  $\omega''$  expires – that this firm earns after the first unsuccessful attempt is

$$(49) \quad A_{22}[q_{j,N_j-1}] = V_{\omega''} \left[ \omega', \omega'', \Omega^-[ \tau_{j'',n''} + L ] \left| (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}] \right. \right].$$

If the first screening was not fruitful, and if  $A_{22}[q_{j,N_j-1}] > 0$ , then this firm will continue the search by screening the remaining compound, namely the compound  $(x_{j,N_j}, q_{j,N_j}, c_{j,N_j})$  immediately, and under this scenario, (49) represents the discounted payoff net of R&D cost – discounted to the time the patent on the brand  $\omega''$  expires – that is yielded by the second screening.

The expected discounted profit net of R&D cost made by the firm which holds the patent on the brand  $\omega''$  – discounted to the time this patent expires – if this firm chooses to screen the penultimate compound immediately after the patent on the brand  $\omega''$  expires is then given by

$$(50) \quad A_{23} = -\gamma_j + \sum_{q_{j,N_j-1} \in Q_{j,N_j-1}[\omega']} A_{21}[q_{j,N_j-1}] f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}] \\ + \left(1 - \sum_{q_{j,N_j-1} \in Q_{j,N_j-1}[\omega']} f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}]\right) A_{22}[q_{j,N_j-1}]$$

Let

$$(51) \quad V_{\omega''} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \left| \left( (x_{j,n}, q_{j,n}, c_{j,n}) \right)_{n=N_j-1}^{N_j}, f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}] \right. \right] \\ = \max\{0, A_{23}\}$$

denote the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds this patent, given that the remaining two compounds are the last two compounds in the  $j$ th chemical library and that the uncertain quality of each of these two compounds is characterized by the common probability mass function  $f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}]$ . The firm which holds the patent on the brand  $\omega''$  will only carry out the screening if this expected payoff is positive.

The information generated from the screening of the penultimate compound in the  $j$ th chemical library resolves not only the quality of this compound, but it also yields information about the potential quality of the last compound. The information thus obtained on the potential quality of the last compound will have a bearing on the decision whether or not this compound will be screened. According to Lemma 1, if the screening of the penultimate compound is not particularly fruitful, then expectations concerning the potential quality of the last compound will be much lowered, and this compound might never be screened, especially when there are many brands competing for a limited amount of drug expenditures. On the other hand, if the screening of the penultimate compound is fruitful, then the potential quality of the last compound will also be stochastically larger. However, this does not necessarily mean that the last compound will be screened for a possible new drug. Indeed, if the

screening of the penultimate compound is particularly fruitful, then no efforts will be expended in screening the last compound, as asserted by the following proposition:

**PROPOSITION 8:** *Suppose that the penultimate compound in the  $j$ th chemical library is screened by the firm which holds the patent on the brand  $\omega''$  when this patent expires. If  $q_{j,N_j-1}$ , the revealed quality of the penultimate compound, is particularly high, then the probability that the quality of the last compound in the  $j$ th chemical library is higher than  $q_{j,N_j-1}$  – so that the drug developed from the last compound can be patented – is particularly low, and this means that the last compound will not be screened.*

**PROOF:** The proof and a numerical example in support of Proposition 8 is given in Annex J.

Let

$$(52) V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \mid ((x_{j,n}, q_{j,n}, c_{j,n}))_{n=N_j-1}^{N_j}, f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}] \right]$$

denote the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$ , given that the remaining two compounds are the last two compounds in the  $j$ th chemical library and that the uncertain quality of each of these two compounds is characterized by the common probability mass function  $f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}]$ . This expected discounted payoff can be computed as follows.

First, if the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  immediately after the patent on the brand  $\omega''$  has expired, and if the screening results in a patentable drug, say  $\omega^1 = (x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1}, \tau_{j,N_j-1})$ , with  $\tau_{j,N_j-1} = \tau_{j'',n''} + L$ , then the expected

discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm that holds the patent on the brand  $\omega'$  is given by

(53)

$$A_{24} = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega'}[s, m, \omega^1, \omega', \Omega^-[s]] ds + e^{-r(\tau_{j',n'}-\tau_{j'',n''})} \times \\ V_{\omega'} \left[ \omega^1, \omega', \Omega^-[\tau_{j',n'} + L] \mid (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}] \right].$$

In (53),  $\Omega^-[\tau_{j',n'} + L] = \Omega^-[\tau_{j'',n''} + L] \cup \{\omega'\}$ .

If the screening of the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  is a failure, then immediately after this screening, we have the situation analyzed in Sub-section 3.4.1, and the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$  is given by

$$(54) \quad A_{25} = V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \mid (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])}[q_{j,N_j}] \right].$$

Thus, under the scenario that the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  immediately after the patent on the brand  $\omega''$  has expired, then the expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – for the firm which holds the patent on the brand  $\omega'$  is given by

$$(55) \quad A_{26} = \sum_{q_{j,N_j-1} \in Q_{j,N_j-1}[\omega']} A_{24} f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}] \\ + \left( 1 - \sum_{q_{j,N_j-1} \in Q_{j,N_j-1}[\omega']} f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}[q_{j,N_j-1}] \right) A_{25}.$$

Second, if the firm which holds the patent on the brand  $\omega''$  chooses not to carry out any screening when the patent on the brand  $\omega''$  expires, then the expected discounted profit – discounted to the time the patent on the brand  $\omega''$  expires – earned by the brand  $\omega'$  from time  $\tau_{j'',n''} + L$  until time  $\tau_{j',n'} + L$  is given by

$$(56) \quad A_{27} = \int_{\tau_{j'',n''}+L}^{\tau_{j',n'}+L} e^{-r(s-\tau_{j'',n''}-L)} v_{\omega'}[s, m, \omega', \Omega^-[s]] ds.$$

Furthermore, the firm which holds the patent on the brand  $\omega'$  is the only one left in the R&D sector. At time  $\tau_{j',n'} + L$ , when the patent on the brand  $\omega'$  expires, this firm might choose not to screen one of the two remaining compounds, and the game ends at time  $\tau_{j',n'} + L$ . On the other hand if it chooses to screen the penultimate compound, and obtains a patentable drug, say  $\omega^1 = (x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1}, \tau_{j,N_j-1})$ , with  $\tau_{j,N_j-1} = \tau_{j',n'} + L$ , then the expected discounted payoff – discounted to time  $\tau_{j',n'} + L$  – obtained by this firm is given by

$$(57) \quad A_{28} = \int_{\tau_{j',n'}+L}^{\tau_{j',n'}+2L} e^{-r(s-\tau_{j',n'}-L)} v_{\omega^1}[s, m, \omega^1, \Omega^-[s]] ds \\ + e^{-rL} V_{\omega^1} \left[ \Omega^-[ \tau_{j',n'} + 2L ] \mid (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}] \right].$$

On the other hand, if the screening of the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1}, \tau_{j,N_j-1})$  is not fruitful, then the expected discounted payoff – discounted to time  $\tau_{j',n'} + L$  – obtained by this firm after the patent on the brand  $\omega'$  has expired is given by

$$(58) \quad A_{29} = V_{\omega'} \left[ \Omega^-[ \tau_{j',n'} + L ] \mid (x_{j,N_j}, q_{j,N_j}, c_{j,N_j}), f_{(\alpha_j[N_j], \beta_j[N_j])} [q_{j,N_j}] \right].$$

The expected discounted payoff – discounted to the time the patent on the brand  $\omega'$  expires – earned by the firm which holds the patent on the brand  $\omega'$  after this patent has expired, given that (i) the firm which holds the patent on the brand  $\omega''$  chooses not to carry out any screening when this patent expires, and (ii) the firm which holds the patent on the brand  $\omega'$  chooses to screen the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  when the patent on the brand  $\omega'$  expires, is then given by

$$(59) \quad A_{30} = -\gamma_j + \sum_{q_{j,N_j-1} \in Q_{j,N_j-1}[\omega']} A_{28} f_{(\alpha_j[N_j-1], \beta_j[N_j-1])} [q_{j,N_j-1}] \\ + \left( 1 - \sum_{q_{j,N_j-1} \in Q_{j,N_j-1}[\omega']} f_{(\alpha_j[N_j-1], \beta_j[N_j-1])} [q_{j,N_j-1}] \right) A_{29}.$$

The expected discounted payoff – discounted to the time the patent on the brand  $\omega''$  expires – earned by the firm which holds the patent on the brand  $\omega'$  after this patent has expired, given that (i) the firm which holds the patent on the brand  $\omega''$  chooses not to carry out any screening when this patent expires, and (ii) the firm which holds the patent on the brand  $\omega'$  chooses to screen the compound  $(x_{j,N_j-1}, q_{j,N_j-1}, c_{j,N_j-1})$  when the patent on the brand  $\omega'$  expires, is then given by

$$(60) \quad A_{31} = A_{27} + e^{-r(\tau_{j',n'} - \tau_{j'',n''})} \max\{0, A_{30}\}.$$

We are now ready to give the explicit expression for (52) as

(61)

$$V_{\omega'} \left[ \omega', \omega'', \Omega^-[\tau_{j'',n''} + L] \mid ((x_{j,n}, q_{j,n}, c_{j,n}))_{n=N_j-1}^{N_j}, f_{(\alpha_j[N_j-1], \beta_j[N_j-1])}(q_{j,N_j-1}) \right]$$

$$= \begin{cases} A_{26} & \text{if the firm which holds the patent on the brand } \omega'' \text{ chooses to screen the} \\ & \text{penultimate compound when the patent on the brand } \omega'' \text{ expires,} \\ A_{31} & \text{if the firm which holds the patent on the brand } \omega'' \text{ chooses not to} \\ & \text{screen the penultimate compound when the patent on the brand } \omega'' \text{ expires.} \end{cases}$$

### 3.4.3 Three or More Remaining Unscreened Chemical Compounds

When there remain three unscreened compounds, they might be located in three different chemical libraries or two of them in the same chemical library and the third one in another chemical library. The analysis carried out in Sub-section 3.4.2 can be used as the basis for the backward induction needed to solve the sub-game in which three unscreened compounds remain. Because of the complexity of the model, a simple recurrent relation between the sub-games does not exist.

As we move backward through time to the root of the game tree, we find fewer and fewer brands on the market and less and less information on the potential qualities of the unscreened compounds. The small number of brands on the market at the beginning means less competition on the product market. However, the small number of compounds screened also means less information is available on the potential

qualities of the unscreened compounds, and this means more risk is involved in searching for a new drug. According to Touilly *et al.* (2002), international pharmaceutical companies now concentrate their R&D efforts not on well-known drugs but on very risky new drugs. This choice might reflect the presumption that the risk involved in carrying out R&D activities in less well-known libraries might be more than compensated for by the market share obtained if the R&D activities are fruitful.

### 3.5 Concluding Remarks

Using a dynamic model of optimal patent design and in the presence of information externalities, the evolution of technological progress in the context of a pharmaceutical industry is studied. The preceding literature on the topic works with only one brand, the brand with the highest quality. As well, the demand is assumed to be completely inelastic. In the conventional models of patent design the role of competitive fringe firms is also discussed implicitly. The model discussed in this research is a continuous in-time dynamic model which provides a rigorous structure for studying the context. It considers several differentiated products, both those whose patents are still in force and those whose patents have already expired, at any point in time. Furthermore the demand for a brand is taken to be a function of income, its price, and the prices of other brands. The interaction of the fringe firm with other patent-holding firms is also explicitly considered under this framework. Unlike the previous literature on the context, the model incorporates both product and process innovation concepts and provides real guidelines to measure the patent breadth. Under this structure, pharmaceutical firms with an active drug discovery program behave strategically in their R&D and in the product markets. Under this structure, three scenarios are discussed: the first scenario refers to the case where all chemical compounds in all chemical libraries are already screened. The second scenario discusses the case where the set of unscreened chemical compounds include only one element. Under this scenario it is shown that the firm which holds the

earlier-expiring patent only chooses to screen the last compound, when the patent it holds expires, if the expected discounted payoff net of R&D costs yielded by this action is positive. The expected discounted payoff net of R&D costs obtained by this firm is then decreasing in the cost of screening; increasing in the cumulative quality discovered in the past screenings of the compounds in the chemical library, and decreasing in the number of past screenings carried out in this chemical library. The payoff is also higher if the marginal cost of the drug manufactured from the last compound is lower, and higher if the qualities (marginal costs) of all the other brands – generic drugs as well as the brands whose patents have not expired – are lower (higher). It is also shown that if the firm which holds the earlier-expiring patent chooses not to screen the last compound when the patent it holds expires, then neither will the firm which holds the patent on the rival brand choose to screen the last compound when this patent expires. The expected discounted payoff earned by the rival brand is shown to be decreasing in the cumulative quality discovered in the chemical library, and increasing in the number of past screenings in the chemical library; increasing (decreasing) in the quality (marginal cost) of the brand and decreasing (increasing) in the qualities (marginal costs) of the generic drugs as well as the brand whose patent has not expired. The third scenario refers to the case where the set of unscreened chemical compounds includes two elements. These elements may be located in two different chemical libraries, or both being located in the same chemical library. Under the first case the analysis suggests that if neither compound is screened by the firm which holds the earlier-expiring patent when it stands alone as the single remaining compound, then it will not be screened either by this firm when it is one of the two remaining compounds. It is also shown that the expected discounted payoff net of R&D costs that the firm which holds the earlier-expiring patent obtains by screening each of the last two compounds is positive, then, every other thing equal, the compound with the lower screening cost should be screened first, the compound from which a new drug with much higher marginal cost is manufactured should not be screened first, and the compound with the potential quality that is stochastically much larger should be screened first. When the elements

are both located in the same chemical library the penultimate compound in the chemical library is screened by the firm which holds the earlier-expiring patent, and if the revealed quality of the penultimate compound, is particularly high, then the probability that the quality of the last compound in the chemical library is higher than the quality of the penultimate compound is particularly low, which means the last compound will not be screened.

This work briefly discusses the concepts of patent length and breadth. An interesting extension to this work would be to use the model's formulation to abstract the optimal values for the breadth and length of patent. The model also provides a rigorous structure for analyzing the context under a strategic timing framework. A possible extension to the current work would then be a discussion that offers insights on the impact of strategic timing and the patent race on the subject.

## **Annex A: The Proof of Proposition 1**

PROOF: A rise in the effective marginal cost  $\bar{c}_{j',n'}$  of the brand  $\omega' = (x_{j',n'}, q_{j',n'}, c_{j',n'}, \tau_{j',n'})$  shifts both curves  $\Phi_{\omega'}$  and  $\Phi_{\omega''}$  upward, inducing a rise in the effective prices of both brands. Because the expenditure on the numéraire good is a constant fraction of  $m[t] + \bar{p}_{\omega'} + \bar{p}_{\omega''} + \bar{c}_{j,n}$ , which is higher at the new equilibrium, total expenditures on drugs must be lower under the new equilibrium.

Using the fact that  $m[t] + \bar{p}_{\omega'} + \bar{p}_{\omega''}$  is higher under the equilibrium, we can assert that the demand for and a fortiori the total revenues earned by the generic drugs must be higher under the new equilibrium, and this means the market share for generic drugs is higher under the equilibrium.

The rise in  $\bar{p}_{\omega'}$ , by raising  $m_{\omega''}[t] = m[t] + \bar{p}_{\omega'} + \bar{c}_{j,n}$ , shifts the inverse demand curve for  $\bar{y}_{\omega''}$  upward by the same proportion at each level of demand, and this means an upward shift in the marginal revenue curve associated with this inverse market demand curve. At the new equilibrium, the output, the total revenue, and the profits earned by the brand  $\omega''$  are all higher. Also, the market share for the brand  $\omega''$  is higher under the new equilibrium.

Using the results just proven that the market share for the brand  $\omega''$  and the market share for generic drugs are both higher under the new equilibrium, we can assert that the market share for the brand  $\omega'$  is lower under the new equilibrium.

Intuitively, we should expect that a higher effective marginal cost of a brand reduces its profitability. However, the technical arguments required to support this intuition are not straightforward. On the one hand, a rise in the effective marginal cost of the brand  $\omega'$ , every other thing equal, reduces the profit earned by this brand. On the other hand, the firm that holds the patent on the rival brand  $\omega''$  behaves strategically by raising the per-therapeutic-unit price of its own brand, and this action induces an upward shift in the demand curve for the brand  $\omega'$ , which has a positive impact on the profitability of the brand  $\omega'$ . Because these two effects are in opposite directions, the net impact of a higher effective marginal cost of the brand  $\omega'$  on its own profitability cannot be determined unambiguously without some efforts. First, let us rewrite (13), the profit earned by this brand, as follows:

$$(A.1) \quad (\bar{p}_{\omega'} - \bar{c}_{j',n'}) \left( \kappa_1 \frac{m_{\omega'}[t]}{\bar{p}_{\omega'}} - \kappa_0 \right) = \left( 1 - \frac{\bar{c}_{j',n'}}{\bar{p}_{\omega'}} \right) \left( \kappa_1 \frac{m_{\omega'}[t]}{\bar{p}_{\omega'}} - \kappa_0 \right) \bar{p}_{\omega'}.$$

Note that (A.1) expresses the profit earned by the brand  $\omega'$  as the product of its total revenue and the factor  $\left( 1 - \frac{\bar{c}_{j',n'}}{\bar{p}_{\omega'}} \right)$ . We have already argued that the total revenue earned by this brand is lower at the new equilibrium. Hence we will succeed in showing that the profit this brand earns will be lower under the new equilibrium if we

can show that the ratio  $\frac{\bar{c}_{j',n'}}{\bar{p}_{\omega'}}$  is higher under the new equilibrium. To this end, recall that the equilibrium effective price of the brand  $\omega'$  is the fixed point of the composite map  $\Phi_{\omega'}$ ; that is,

$$(A.2) \quad \bar{p}_{\omega'} = \frac{1}{\sqrt{\kappa_0}} \sqrt{\kappa_1} \sqrt{\bar{c}_{j',n'}} \sqrt{m[t] + \bar{c}_{j,n} + \frac{\sqrt{\kappa_1} \sqrt{m[t] + \bar{p}_{\omega'} + \bar{c}_{j,n}} \sqrt{\bar{c}_{j'',n''}}}{\sqrt{\kappa_0}}},$$

Squaring (A.2), and then rearranging the result, we obtain

$$(A.3) \quad \frac{\bar{p}_{\omega'}}{\bar{c}_{j',n'}} = \frac{1}{\bar{p}_{\omega'} \sqrt{\kappa_0}} \sqrt{\kappa_1} \sqrt{\bar{c}_{j',n'}} \sqrt{m[t] + \bar{c}_{j,n} + \frac{\sqrt{\kappa_1} \sqrt{m[t] + \bar{p}_{\omega'} + \bar{c}_{j,n}} \sqrt{\bar{c}_{j'',n''}}}{\sqrt{\kappa_0}}}.$$

It can be seen immediately that the expression on the right-hand side of (20) is strictly decreasing in  $\bar{p}_{\omega'}$ . Hence the ratio  $\frac{\bar{c}_{j',n'}}{\bar{p}_{\omega'}}$  will be higher under the new equilibrium, as desired. ■

## Annex B: The Proof of Proposition 2

PROOF: A rise in the effective marginal cost  $\bar{c}_{j,n}$  of the generic drugs shifts both curves  $\Phi_{\omega'}$  and  $\Phi_{\omega''}$  upward, inducing a rise in the per-therapeutic-unit prices of both brands. Because the expenditure on the numéraire good is a constant fraction of  $m[t] + \bar{p}_{\omega'} + \bar{p}_{\omega''} + \bar{c}_{j,n}$ , which is higher at the new equilibrium, total expenditures on drugs must be lower under the new equilibrium.

The rise in  $\bar{p}_{\omega'}$  and  $\bar{c}_{j,n}$  raises  $m_{\omega''}[t]$ , which shifts the inverse demand curve for  $\bar{y}_{\omega''}$  upward by the same proportion at each level of demand, and this means an upward shift in the marginal revenue curve associated with this inverse market demand curve. At the new equilibrium, the output, the total revenue, and the profits earned by the brand  $\omega''$  are all higher. Also, because the total expenditures on drugs are lower

under the new equilibrium, the market share for the brand  $\omega''$  is also higher. The results just established for the brand  $\omega''$  also hold for the brand  $\omega'$ .

At the new equilibrium the lower total expenditures on drugs coupled with the higher expenditure on each of the brands  $\omega'$  and  $\omega''$  imply a lower level of total revenues earned by generic drugs. Also, because the market share for the brand  $\omega'$  and the market share for the brand  $\omega''$  are both higher under the new equilibrium, the market share for generic drugs must be lower under the new equilibrium. ■

### **Annex C: The Proof of Proposition 3**

PROOF: Statement (i) is obvious. To prove (ii), first note that  $A_2[q_{j,N_j}]$  is increasing in  $q_{j,n}$ , and then apply Lemma 1, which asserts that the potential quality of the last compound is stochastically increasing in  $\alpha_j[N_j]$  and  $\beta_j[N_j]$ . Statement (iii) follows immediately from the properties of  $A_1[q_{j,N_j}]$ . ■

### **Annex D: The Proof of Lemma 2**

PROOF: If the firm which holds the patent on the brand  $\omega''$  chooses to screen the last compound when the patent it holds on the brand  $\omega''$  expires, then at each age below  $\tau_{j',n'} - \tau_{j'',n''}$ , during its statutory life, the newly discovered drug has to compete against the brand  $\omega'$  whose patent has not expired and which is sold above its marginal cost, as well as against the generic drugs in  $\Omega^-[ \tau_{j'',n''} + L ]$ . After that and until the end of its statutory life, the newly discovered drug has to compete against all the generic drugs in  $\Omega^-[ \tau_{j',n'} + L ] = \Omega^-[ \tau_{j'',n''} + L ] \cup \{\omega'\}$ .

Now suppose that  $A_3 \leq 0$ ; that is, the firm which holds the patent on the brand  $\omega''$  chooses not to screen the last compound. If the rival firm, which holds the patent on the brand  $\omega'$  chooses to screen the last compound when the patent it holds on the

brand  $\omega'$  expires, and if the screening leads to a patentable drug, then at each age during its statutory life the newly discovered drug has to compete against all the generic drugs in  $\Omega^-[ \tau_{j',n'} + L ] = \Omega^-[ \tau_{j'',n''} + L ] \cup \{\omega'\}$ . Thus, at each age during its statutory life, the patent on the newly discovered drug allows its holder to earn more or the same profit under the scenario that it is discovered by the firm which holds the patent on the brand  $\omega''$  than under the scenario that it is discovered by the firm which holds the patent on the brand  $\omega'$ . ■

## Annex E: The Proof of Proposition 4

PROOF: If the firm which holds the patent on the brand  $\omega''$  chooses not to screen the last compound when the patent on the brand  $\omega''$  expires, then the discounted payoff for the firm which holds the patent on the brand  $\omega'$  is given by  $A_5$ . If the firm which holds the patent on the brand  $\omega''$  chooses to screen the last compound when the patent on the brand  $\omega''$  expires, then the discounted payoff for the firm which holds the patent on the brand  $\omega'$  is still given by  $A_5$  if the screening is not fruitful. On the other hand, if the screening yields a patentable drug, then the discounted payoff for the firm which holds the patent on the brand  $\omega'$  is given by  $A_4[q_{j,N_j}]$ , which is lower than  $A_5$ , and which is decreasing in the quality of the newly discovered drug. Invoking Lemma 1, we can then assert that the discounted payoff for the firm which holds the patent on the brand  $\omega'$  is lower when the potential quality of the last compound is stochastically larger. That is, when the cumulative quality discovered in the *jth chemical library* is higher and when the size of this chemical library is smaller, the expected discounted payoff obtained by the firm which holds the patent on the brand  $\omega'$  will be lower. The remaining statements of Proposition 5 can be proved by applying Propositions 1 and 2. ■

## Annex F: The Proof of Lemma 3

PROOF: (i) Suppose that the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  first.

If the screening is fruitful, then during the time interval  $[\tau_{j'',n''} + L, \tau_{j',n'} + L)$  the newly discovered drug competes against the same set of brands – the generic drugs in  $\Omega^-[\tau_{j'',n''} + L]$  and the brand  $\omega'$  whose patent is still in force – as if it were discovered under the hypothesis that it stands alone as the last remaining compound. At time  $\tau_{j',n'} + L$ , when the patent on the brand  $\omega'$  expires, the firm which holds the patent on this brand will choose not to screen the compound  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$ . Indeed, if the firm which holds the patent on the brand  $\omega'$  chooses to screen the compound  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$  when the patent it holds on the brand  $\omega'$  expires and if the screening is fruitful, then the new drug it discovered will face competition from the drug manufactured from the brand  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  – both before and after the patent on this brand expires – as well as from the generic drugs in  $\Omega^-[\tau_{j'',n''} + L] \cup \{\omega'\}$ . Such an action will yield an expected discounted payoff net of R&D costs that is lower than  $A_7^{\ell'} \leq 0$ , which is clearly not profitable. Thus, if the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  first, then the expected discounted payoff net of R&D costs that it obtains will also be given by  $A_7^{\ell} \leq 0$ . In the same manner, we can show that the firm which holds the patent on the brand  $\omega''$  will not choose to screen the compound  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$  first when the patent it holds on the brand  $\omega''$  expires. We have just shown that when  $A_7^{\ell} \leq 0, \ell = 1, 2$ , the firm which holds the patent on the brand  $\omega''$  will shut down its R&D activities and exit the market after this patent has expired.

Invoking Lemma 2, and using the assumption  $A_7^{\ell} \leq 0, \ell = 1, 2$ , we can assert that the firm which holds the patent on the brand  $\omega'$  will not choose to screen either

compound – if it stands alone as the single remaining compound – when the patent it holds on the brand  $\omega'$  expires. The argument used for the firm which holds the patent on the brand  $\omega''$  can be repeated to assert that the firm which holds the patent on the brand  $\omega'$  will also shut down its R&D activities and exit the market when the patent it holds on the brand  $\omega'$  expires.

(ii) Suppose that  $A_7^\ell > 0, A_7^{\ell'} \leq 0$ . We have argued that if the firm which holds the patent on the brand  $\omega''$  screens the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  first, then the firm which holds the patent on the brand  $\omega'$  will shut down its R&D activities and exit the market when the latter patent expires. This action yields an expected discounted payoff net of R&D costs  $A_7^\ell > 0$  for the firm which holds the patent on the brand  $\omega''$ . Furthermore, because  $A_7^{\ell'} \leq 0$ , when the patent on the new drug manufactured from the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  expires, the firm which holds the patent on the drug manufactured from this compound will not choose to screen the remaining compound because such an action will not yield a positive expected discounted payoff net of R&D costs. Thus the expected discounted payoff net of R&D costs that the firm which holds the patent on the brand  $\omega''$  obtains by screening the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  first will be equal to  $A_7^\ell > 0$ .

If the firm which holds the patent on the brand  $\omega''$  chooses to screen the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  first, and if the screening is fruitful, then the newly discovered drug faces competition from same group of competing brands which exist under the scenario that the preceding compound stands alone as the single remaining compound, not as one of the two remaining compounds, plus possibly the potential competition from another new drug that the firm which holds the patent on the brand  $\omega'$  might develop if the latter firm chooses to screen the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$ . Thus, the expected discounted payoff net of R&D costs that is yielded by screening the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  first is

not positive, and the firm which holds the patent on the brand  $\omega''$  will choose to screen the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  immediately after the patent it holds on the brand  $\omega''$  expires. Furthermore, the firm which holds the patent on the brand  $\omega'$  will not choose to screen the compound  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$  when the patent it holds on the brand  $\omega'$  expires. This firm will shut down its R&D activities and exit the market at time  $\tau_{j',n'} + L$ , and so will the firm which holds the patent on the brand  $\omega''$  once the patent on the drug it develops from the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$  expires. ■

## Annex G: The Proof of Proposition 5

PROOF: Let us rewrite  $A_{11}^\ell$  as follows:

(G.1)

$$\begin{aligned}
A_{11}^\ell &= -\gamma_{j[\ell]} + \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} A_{10}^\ell[q_{j[\ell],N_{j[\ell]}}] f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}] \\
&\quad + \left(1 - \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}]\right) \left((A_7^{\ell'} + \gamma_{j[\ell']}) - \gamma_{j[\ell']}\right) \\
&= \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} A_{10}^\ell[q_{j[\ell],N_{j[\ell]}}] f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}] \\
&\quad + \left(1 - \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}]\right) (A_7^{\ell'} + \gamma_{j[\ell']}) \\
&\quad - \left(\gamma_{j[\ell]} + \left(1 - \sum_{q_{j[\ell],N_{j[\ell]}} \in Q_{j[\ell],N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell],N_{j[\ell]}}]\right) \gamma_{j[\ell']}\right) \\
&= H_1^\ell + H_2^\ell - H_3^\ell.
\end{aligned}$$

Note that  $A_7^{\ell'} + \gamma_{j[\ell']}$  represents the discounted profits gross of R&D costs earned by the firm which holds the patent on the brand  $\omega''$  from the screening of the compound  $(x_{j[\ell'],N_{j[\ell']}}, q_{j[\ell'],N_{j[\ell']}}, c_{j[\ell'],N_{j[\ell']}})$  after the unfruitful screening of the compound  $(x_{j[\ell],N_{j[\ell]}}, q_{j[\ell],N_{j[\ell]}}, c_{j[\ell],N_{j[\ell]}})$ . Also, recall that

$$A_{10}^{\ell} [q_{j[\ell], N_{j[\ell]}}] = \int_{\tau_{j'', n''} + L}^{\tau_{j', n'} + L} e^{-r(s - \tau_{j'', n''} - L)} v_{\omega^{\ell}} [s, m, \omega^{\ell}, \omega', \Omega^{-}[s]] ds \\ + e^{-r(\tau_{j', n'} - \tau_{j'', n''})} A_9^{\ell} [q_{j[\ell], N_{j[\ell]}}].$$

Note that the higher is the cost of screening the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$ , the lower will be the incentive for the firm which holds the patent on the brand  $\omega'$  to screen this compound in which process might discover a new drug that could compete against the drug discovered by screening the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$ . Thus,  $A_9^{\ell} [q_{j[\ell], N_{j[\ell]}}]$  will be higher, the higher is the cost of screening the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$ .

Suppose that  $c_{j[\ell], N_{j[\ell]}} = c_{j[\ell'], N_{j[\ell]}}$  and  $f_{(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}])} = f_{(\alpha_{j[\ell']}[N_{j[\ell]}], \beta_{j[\ell']}[N_{j[\ell]}])}$ , i.e., the marginal costs of the drugs manufactured from the two remaining compounds are the same, and the probability mass functions that characterize the potential qualities of the two remaining compounds are identical. If  $\gamma_{j[\ell]} < \gamma_{j[\ell']}$ , then  $A_{10}^{\ell} [q_{j[\ell], N_{j[\ell]}}] > A_{10}^{\ell'} [q_{j[\ell'], N_{j[\ell]}}]$ , which implies  $H_1^{\ell} > H_1^{\ell'}$ , and  $A_7^{\ell'} + \gamma_{j[\ell']} = A_7^{\ell} + \gamma_{j[\ell]}$ , which implies  $H_2^{\ell} = H_2^{\ell'}$ . Also, if  $\gamma_{j[\ell]} < \gamma_{j[\ell']}$ , then

$$-H_3^{\ell} + H_3^{\ell'} = (\gamma_{j[\ell']} - \gamma_{j[\ell]}) \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} f_{(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}])} [q_{j[\ell], N_{j[\ell]}}] > 0.$$

Hence if  $\gamma_{j[\ell]} < \gamma_{j[\ell']}$ , then  $H_1^{\ell} - H_1^{\ell'} - H_3^{\ell} + H_3^{\ell'} > 0 \Leftrightarrow A_{11}^{\ell} > A_{11}^{\ell'}$ ; that is, every other thing equal, the compound with the lower screening cost should be screened first. ■

## Annex H: The Proof of Proposition 6

PROOF: Suppose that  $\gamma_{j[\ell]} = \gamma_{j[\ell']}$  and  $f_{(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}])} = f_{(\alpha_{j[\ell']}[N_{j[\ell]}], \beta_{j[\ell']}[N_{j[\ell]}])}$ , i.e., the screening costs and the probability mass functions that characterize the

potential qualities of the two remaining compounds are identical. Also, suppose that  $c_{j[\ell'], N_{j[\ell']}} \gg c_{j[\ell], N_{j[\ell]}}$ , i.e., the marginal cost of the drug manufactured from the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  is much higher than the marginal cost of the drug manufactured from the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$ . The differential expected discounted payoff between the action of screening the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  first and the action of screening the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  first is given by

$$\begin{aligned}
\text{(H.1)} \quad A_{11}^{\ell} - A_{11}^{\ell'} = & -\gamma_{j[\ell]} + \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}} [\omega'] A_{10}^{\ell} [q_{j[\ell], N_{j[\ell]}}] f(\alpha_{j[\ell]} [N_{j[\ell]}], \beta_{j[\ell]} [N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}}] \\
& - \left( -\gamma_{j[\ell']} + \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}} [\omega'] A_{10}^{\ell'} [q_{j[\ell'], N_{j[\ell']}}] f(\alpha_{j[\ell']} [N_{j[\ell']}], \beta_{j[\ell']} [N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] \right) \\
& + \left( 1 - \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}} [\omega'] f(\alpha_{j[\ell]} [N_{j[\ell]}], \beta_{j[\ell]} [N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}}] \right) A_7^{\ell'} \\
& - \left( 1 - \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}} [\omega'] f(\alpha_{j[\ell']} [N_{j[\ell']}], \beta_{j[\ell']} [N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] \right) A_7^{\ell} \\
& = H_4^{\ell} - H_4^{\ell'} + H_5^{\ell'} - H_5^{\ell};
\end{aligned}$$

Now if  $c_{j[\ell'], N_{j[\ell']}}$  is high enough so that the discounted profits earned from the new drug manufactured from the brand  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  are not sufficient to justify the R&D costs, then  $A_7^{\ell'} = 0$ . When  $c_{j[\ell'], N_{j[\ell']}}$  is high, but not prohibitive,  $A_7^{\ell'} > 0$ , but not too high. Under this scenario, if the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  is screened first, and if the screening is fruitful, then the drug manufactured from this compound will not face much competition from the drug manufactured from the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$ . In this case,  $H_4^{\ell} \cong A_7^{\ell}$  and  $H_4^{\ell'} < A_7^{\ell'}$ , and the differential expected payoff can be approximated by

$$\begin{aligned}
\text{(H.2)} \quad A_{11}^\ell - A_{11}^{\ell'} &\geq \\
(A_7^\ell - A_7^{\ell'}) - \left(1 - \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}]}\right) (A_7^\ell - A_7^{\ell'}) \\
&= \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}]} (A_7^\ell - A_7^{\ell'}) > 0,
\end{aligned}$$

which implies that the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  should be screened first. ■

## Annex I: The Proof of Proposition 7

PROOF: The expected discounted payoff differential between screening the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  first and screening the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$  first is given by

$$\begin{aligned}
\text{(I.1)} \quad A_{11}^\ell - A_{11}^{\ell'} &= \\
&-\gamma_{j[\ell]} + \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} A_{10}^\ell [q_{j[\ell], N_{j[\ell]}]} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}]}] \\
&+ \left(1 - \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}]}\right) A_7^{\ell'} \\
&-\left(-\gamma_{j[\ell']} + \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}[\omega']} A_{10}^{\ell'} [q_{j[\ell'], N_{j[\ell']}}] f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] \right) \\
&\quad - \left(1 - \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}[\omega']} f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] \right) A_7^\ell.
\end{aligned}$$

If we let

$$\begin{aligned}
\text{(I.2)} \quad H_6^\ell &= -\gamma_{j[\ell]} + \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} A_{10}^\ell [q_{j[\ell], N_{j[\ell]}]} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}]}], \\
&(\ell = 1, 2),
\end{aligned}$$

then we can rewrite (38) as

(I.3)

$$\begin{aligned} \frac{A_{11}^\ell - A_{11}^{\ell'}}{H_6^\ell} = 1 + & \left( 1 - \sum_{q_{j[\ell], N_{j[\ell]}} \in Q_{j[\ell], N_{j[\ell]}}[\omega']} f(\alpha_{j[\ell]}[N_{j[\ell]}], \beta_{j[\ell]}[N_{j[\ell]}]) [q_{j[\ell], N_{j[\ell]}}] \right) \frac{A_7^{\ell'}}{H_6^\ell} - \frac{H_6^{\ell'}}{H_6^\ell} \\ & - \left( 1 - \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}[\omega']} f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] \right) \frac{A_7^\ell}{H_6^\ell}. \end{aligned}$$

If the potential quality of the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  is stochastically much larger than the potential quality of the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$ , then  $A_7^\ell$  is much larger than  $A_7^{\ell'}$ , i.e.,  $\frac{A_7^{\ell'}}{A_7^\ell}$  is small. Furthermore, when the former compound is screened first, there is a greater chance of obtaining a drug of high quality from this action, and thus less chance that it will face competition from a possible new drug that might be discovered by the firm which holds the patent on the brand  $\omega'$  when this patent expires, and this means  $H_6^\ell \cong A_7^\ell$ . Also,  $\frac{H_6^{\ell'}}{H_6^\ell}$  is small because  $H_6^{\ell'} \leq A_7^{\ell'}$ .

Thus, when the potential quality of the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  is stochastically much larger than that of the compound  $(x_{j[\ell'], N_{j[\ell']}}, q_{j[\ell'], N_{j[\ell']}}, c_{j[\ell'], N_{j[\ell']}})$ , we have

(I.4)

$$\begin{aligned} \frac{A_{11}^\ell - A_{11}^{\ell'}}{H_6^\ell} & \cong 1 - \left( 1 - \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}[\omega']} f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] \right) \\ & = \sum_{q_{j[\ell'], N_{j[\ell']}} \in Q_{j[\ell'], N_{j[\ell']}}[\omega']} f(\alpha_{j[\ell']}[N_{j[\ell']}], \beta_{j[\ell']}[N_{j[\ell']}]) [q_{j[\ell'], N_{j[\ell']}}] > 0, \end{aligned}$$

and this means that the compound  $(x_{j[\ell], N_{j[\ell]}}, q_{j[\ell], N_{j[\ell]}}, c_{j[\ell], N_{j[\ell]}})$  should be screened first. ■

## Annex J: The Proof of Proposition 8

PROOF: The probability mass function that characterizes the potential quality of the last compound in the  $j$ th chemical library, after the quality of the penultimate compound has been revealed, is  $f_{(\alpha_j[N_j], \beta_j[N_j])}(q_{j,N_j})$ , where  $\alpha_j[N_j] = \alpha_j[N_j - 1] + q_{j,N_j-1}$  and  $\beta_j[N_j] = \frac{1}{1 + \frac{1}{\beta_j[N_j-1]}}$ . The probability that the quality of the last compound exceeds  $q_{j,N_j-1}$  is

(J.1)

$$\text{Prob} \{q_{j,N_j} > q_{j,N_j-1} \mid q_{j,N_j-1}\} = \int_0^\infty \left( \sum_{q > q_{j,N_j-1}} e^{-\lambda} \frac{\lambda^q}{q!} \right) g_{(\alpha_j[N_j-1] + q_{j,N_j-1}, \beta_j[N_j])}[\lambda] d\lambda.$$

With a heavy dose of limiting arguments, we can show that  $\text{Prob} \{q_{j,N_j} > q_{j,N_j-1} \mid q_{j,N_j-1}\} \rightarrow 0$  when  $q_{j,N_j-1} \rightarrow \infty$ . However, we eschew these technical arguments and offer a numerical example to illustrate this result. For the simple numerical example, suppose that there are three brands – one generic drug and two leading brands – on the market. The quality of the generic drug is 3. As usual the two leading brands, whose patents have not expired, are denoted by  $\omega''$  and  $\omega'$ , respectively, with the patent of the former brand being granted before the patent on the latter. Also, for simplicity suppose that the brand  $\omega''$  is obtained from screening the  $(N_j - 2)$ th compound in the  $j$ th chemical library and that  $N_j = 3$ . The quality of the brand  $\omega''$  is assumed to be 4. As for the brand  $\omega'$ , it is obtained from a different chemical library, and its quality is assumed to be 5. The following table presents the results of the numerical exercise.

**Table H.1. The Probability that the Last Compound in a Chemical Library will be Screened as a Function of the Revealed Quality of the Penultimate Compound**

| $q_{j,N_j-1}$ (the revealed quality of the penultimate compound) | $Prob \{q_{j,N_j} > q_{j,N_j-1}   q_{j,N_j-1}\}$ (the probability that the drug developed from the last compound meets the leading breadth requirement for a patent) |
|------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6                                                                | 0.263                                                                                                                                                                |
| 7                                                                | 0.223                                                                                                                                                                |
| 8                                                                | 0.191                                                                                                                                                                |
| 9                                                                | 0.163                                                                                                                                                                |
| 10                                                               | 0.140                                                                                                                                                                |
| 11                                                               | 0.121                                                                                                                                                                |
| 12                                                               | 0.104                                                                                                                                                                |
| ...                                                              | ...                                                                                                                                                                  |
| 20                                                               | 0.033                                                                                                                                                                |
| 21                                                               | 0.029                                                                                                                                                                |
| 22                                                               | 0.025                                                                                                                                                                |

As can be seen from the preceding table, the probability that the drug manufactured from the last compound meets the leading breadth requirement for a patent declines steadily as the revealed quality of the penultimate compound rises. When the outcome of the screening of the penultimate compound is particularly fruitful, this probability is so low to render the expected discounted profit yielded by the potential new drug developed from the last compound insufficient to cover its R&D costs. ■

## References

- Arrow, K. and S. Chang (1982)**, “Optimal Pricing, Use, and Exploration of Uncertain Resource Stocks,” *Journal of Environmental Economics and Management*, 9, 1-10.
- Cairns, R. and N.V. Quyen (1998)**, “Optimal Exploration for and Exploitation of Heterogenous Mineral Deposits,” *Journal of Environmental Economics and Management*, 35, 164-189.
- Chow, G.C. (1981)**, *Economic Analysis by Control Methods*, New York: John Wiley & Sons.
- Eaton, J. and S. Khortum (2001)**, “Technology, Trade, and Growth: A Unified Framework,” *European Economic Review*, 451, 742-755.
- Gilbert, R. and C. Shapiro (1990)**, “Optimal Patent Length and Breadth,” *The Rand Journal of Economics*, 21, 106-112.
- Klemperer, P. (1990)**, “How Broad Should the Scope of Patent be?” *The Rand Journal of Economics*, 21, 113-130.
- Mansfield, E. (1961)**, “Technical Change and the Rate of Imitation,” *Econometrica*, 29, 741-766.
- **(1984)**, “R&D and Innovation: Some Empirical Findings,” in Zvi Griliches, ed., *R&D, Patents and Productivity*, Chicago: University of Chicago Press for the National Bureau of Economic Research.
- Nordhaus, W. (1969)**, *Invention, Growth, and Welfare*, Cambridge, Massachusetts, MIT Press.
- **(1972)**, “The Optimal Life of the Patent: Reply,” *American Economic Review*, 62, 428-431.
- O’Donoghue, T., S. Scotchmer, and J. Thisse (1998)**, “Patent Breadth, Patent Life, and the Pace of Technological Progress,” *Journal of Economics & Management Strategy*, 7(1), 1-32.
- Pearlman, R. S. (2009)**, “Novel Software Tools for Addressing Chemical Diversity,” 1995-2009 Network Science Corporation.
- Quyen, N. V. (1991)**, “Exhaustible Resources: A Theory of Exploration,” *Review of Economic Studies*, 58, 777-789.
- Scotchmer, S. and J. Green (1990)**, “Novelty and Disclosure in Patent Law,” *The Rand Journal of Economics*, 21, 131-146.

- Takalo, T. (2001)**, "On the optimal patent policy," Finnish Economic Papers, Finnish Economic Association, 14(1), 33-40.
- Tandon, P. (1982)**, "Optimal patents with compulsory licensing, "Journal of Political Economy, 90, 470-486.
- Touilly, V. (2002)**, *Pharmaceutical Industry: Innovation and Economic Performance: A Study on Behalf of Laboratoires Internationaux de Recherche*, 18 pages