Microeconomic Foundations of Knowledge-Driven Growth:

Modelling the Dynamic Allocation of R&D Resources

John Gabriel Goddard
Hertford College, Oxford

Thesis submitted to the Department of Economics, University of Oxford, in fulfilment of the requirements for the degree of Doctor of Philosophy

Michaelmas Term, 2003 [i.e. 2004]

John Gabriel Goddard
Hertford College, Oxford

Doctor of Philosophy
Michaelmas Term, 2003

Microeconomic Foundations of Knowledge-Driven Growth: Modelling the Dynamic Allocation of R&D Resources

Abstract

This D.Phil. thesis undertakes a theoretical analysis of the microeconomic incentives for scientific and technical knowledge-creating activities at the firm-level, the channels by which these activities impinge on industrial change and economic growth, and the effectiveness of governmental policies formulated to influence these systemic linkages. The motivation for this work is explained in Chapter 2, which reviews the state of the art in new growth theory and puts forward a typology of privately sponsored R&D activities and knowledge resources defining the premises on which the thesis rests.

Chapters 3 and 4 investigate the R&D allocation and output decisions of a profit-maximising monopolist investing in exploratory- and applications-oriented research, dealing separately with product and process innovations. The characteristic properties of the optimal time paths are ascertained by means of formal and numerical optimal control methods, including comparative dynamics. The complementarity between the two modes of research is shown to generate increasing returns, but these turn out to be short-lived. The model is extended in Chapter 5 to study the development of multiple product lines. Knowledge spillovers and demand-side externalities across successive product lines can provide the basis for continued spending on R&D, allowing sustained output growth and profitability.

Chapters 6 to 8 turn to the challenges of modelling the irreducible elements of uncertainty in the innovation process and their bearing upon the dynamics of market competition and industry structure. In the sequential game theoretic model introduced, firms can invest in fundamentally uncertain "innovative-R&D," or wait until the uncertainty surrounding original innovation is dispelled and invest in certain but costly "imitative-R&D." These decisions are taken in a vertically and horizontally differentiated market where non-innovating firms can compete with a "traditional" product. The industry-wide scale of R&D investments and the related evolution in market structure are determined endogenously. To do so, a symmetric equilibrium concept is defined and its uniqueness established.

The model can support Schumpeterian industry evolutions, in which surges of innovative entry are followed by waves of imitation, and ensuing "creative destruction" in which traditional producers are driven out of the industry and innovators' rents are eventually eroded. Numerical simulations are employed in Chapter 7 to provide further insights into the evolution of product development, market structure, pricing, firm growth, profitability, and consumer welfare. The final chapter considers the implications of this game theoretic approach for competition and innovation policies.

Acknowledgements

This thesis was worked out under the direction of Professor Paul A. David, to whom I owe a debt that I can only gratefully acknowledge. It was a privilege and always a pleasure to be supervised by Paul David, who imparted his judicious directions and generous advice with great candour and bonhomie and was enormously supportive throughout the doctoral project.

I also must express my thanks to the following people for their suggestions, comments, and criticisms on various chapters of the dissertation: Timothy Bresnahan, Rodrigo Carcamo, Jean-Michel Dalle, Paul Geroski, Bronwyn Hall, Max Keilbach, Franco Malerba, Margaret Meyer, David Myatt, Jose Carlos Ramirez, Ericka Rascon, Suzanne Scotchmer, Emiliano Torres, Justin Vandeven, Linus Wilson, and Murat Yildizoglu.

My postgraduate studies were generously funded by a scholarship (number 119622) from the Consejo Nacional de Ciencia y Tecnologia (Conacyt) in Mexico. I'm obliged to the Centro de Investigacion y Docencia Economicas (CIDE) for its support. Within these institutions, I owe special thanks to Carlos Bazdresch Parada and Carlos Elizondo Mayer-Serra for their steady encouragement.

Hertford College provided a stimulating environment in which to study as well as the opportunity to meet many interesting people. The College directly supported my research through the provision of two graduate travel grants. I am also indebted to the George Webb Medley Endowment Fund and the European Summer School on Industrial Dynamics for further financial support.

My deepest gratitude goes to my parents, sister, and grandparents, who have always been there for me. The same thanks go to the rest of my family and all of my friends, who remain unnamed to avoid missing someone.

Contents

Acknowledgements iii

Contents iv

Chapter 1 Introduction 1

Chapter 2 Abstraction and idealization in the 'mirror of reality': A selective critique of the new growth literature 10

2.1 Introduction
2.2 An examination of the 'state of the art' in new growth theory
  2.2.1 An overview of Aghion and Howitt's 'creative destruction' model
  2.2.2 R&D and knowledge, in theory and in practice
  2.2.3 The supply of and demand for research workers
  2.2.4 Financial markets and innovation: Who wants to "fund a revolution"?
  2.2.5 To be or not to be Schumpeterian? What does "creative destruction" involve?
2.3 Stepping back from the world's complexity – a fresh approach to modelling

Chapter 3 A dynamic model of exploratory- and applications-oriented research allocation by a monopoly enterprise 70

3.1 Introduction
3.2 Notation and Assumptions
3.3 Product quality innovation
3.4 Own-process innovation
3.5 Example with specific functional forms
3.6 Conclusions

Chapter 4 Firm-level R&D Expenditures under Monopoly Conditions: A Numerical Analysis 114

4.1 Introduction
4.2 Outline of the dynamic model
4.3 Numerical methodology
  4.3.1 Benchmark specification
  4.3.2 Boundary value solver
  4.3.3 Sensitivity analysis
4.4 Central results in the benchmark simulations
4.5 Comparative dynamics analysis
  4.5.1 Product market conditions
  4.5.2 Production and research capabilities
  4.5.3 Labour market conditions
  4.5.4 Discount rate
4.6 Innovation in a volatile environment
4.7 Conclusions

Chapter 5 Managing R&D and production across multiple product lines 174

5.1 Introduction
5.2 Model and methodology
  5.2.1 Theoretical framework
  5.2.2 Numerical continuation methods
  5.2.3 Geometric envelope analysis
5.3 Development of multiple product lines with exogenously given transitions
5.4 Optimal number of product lines marketed in a fixed time frame
5.5 The financial feasibility of "accelerated product development"
5.6 Conclusions

Chapter 6 Uncertainty, innovation, imitation and the evolution of market structure 213

6.1 Introduction
6.2 Industrial dynamics under Knightian uncertainty: A formal model
  6.2.1 Notation and assumptions
  6.2.2 The equilibrium concept and a solution procedure
6.3 Analysis of the symmetric sustainable equilibrium
6.4 Conclusions

Chapter 7 Dynamics of industry structure and market competition under uncertainty: A numerical analysis 278

7.1 Introduction
7.2 Numerical solution
  7.2.1 Methodology
  7.2.2 Product development and market structure
  7.2.3 Firm-level trends in output and profitability
  7.2.4 Industry-level outcomes
7.3 Fundamental uncertainty and structural change: A closer inspection
7.4 Conclusions

Chapter 8 The potential for welfare-enhancing competition and innovation policies 311

8.1 Introduction
8.2 Competition, structural change and the public interest
  8.2.1 A comparative analysis of Cournot and Bertrand competition
  8.2.2 Market fragmentation: From niche markets to perfect substitutability
  8.2.3 Implications for competition policy
8.3 The scope for welfare-improving innovation policies
8.4 Towards a policy-relevant research agenda

References 356

Chapter 1

Introduction

This D.Phil. thesis undertakes a theoretical analysis of the microeconomic incentives for scientific and technical knowledge-creating activities at the firm-level, the channels by which these activities impinge on industrial change and economic growth, and the effectiveness of governmental policies formulated to influence these systemic linkages. Examining the implications of strategy choices under a richer set of specifications for the research activities of profit-maximising firms is of interest in its own right. There is a further potential benefit, in that the resulting models may complement existing analyses of innovation in the new growth theory literature, which until now have been based on a number of simplifying assumptions about the 'structuring' and management of research activities. In particular, most extant models of knowledge-driven growth ignore the fundamentally stochastic nature of research outcomes, and do not distinguish among the diverse strategic goals that can be pursued by investment in "R&D." Moreover, they treat R&D investments as homogenous, abstracting from the different classes of knowledge relating to science and technology that can be generated by such expenditures. This is reflected in the reliance of many such models upon a "stock of knowledge," thereby conflating heterogeneous elements of knowledge which have different instrumental value for different firms. This critique of the literature is developed in Chapter 2.

The dissertation's original contributions rest on different premises: research activities and the intangible and tangible knowledge ensuing from them should be explicitly modelled and, furthermore, it is quite feasible to construct an appropriate dynamic framework that will

yield a variety of solvable models with more realistic and precisely delineated microeconomic foundations. The typology of privately sponsored R&D activities and knowledge resources introduced in Chapter 2 identifies several distinct kinds of research activities that are interesting from an economic viewpoint, including: exploratory vis-à-vis applications-oriented research; the partition of the latter into product and process R&D; and the difference between own-product and -process improving research versus that used to reveal and replicate the technical knowledge embodied in other firms' innovative products. This reading of a firm's research activities disqualifies any one-dimensional measure of a knowledge stock as inappropriate for dynamic analysis. It is necessary, instead, to keep track of several indicators of the 'fundamental' and 'practical' knowledge frontiers reached by the firm through the efforts of its researchers; it is, correspondingly, crucial to track the acquisition of new knowledge of different kinds from external sources.

Chapter 3 sets up a dynamic model of exploratory- and applications-oriented research allocation by a monopoly enterprise as a constrained optimal control problem. The assumption of a product monopoly puts aside strategic considerations in order to focus attention on the way a firm phases its R&D programme over time, given the dynamic connections between the heterogeneous research activities and production. Exploratory research, broadly conceived of as the search for an understanding of general phenomena, and the intangible knowledge that can be embodied in research platforms or research tools, is treated as contributing to the firm's profits only indirectly. The findings do not generate an immediate flow of revenue, but instead may be used by applications-oriented researchers labouring to improve product designs or the production technology. A comparison of the results for a model in which the firm's exploratory research is complementary to new product-oriented R&D, or to 'new' process R&D, reveals that the central difference between the two is reflected in the optimised production plan. The optimal research plan implies that

the firm will gain fundamental knowledge resources at a slowing pace in any specific research domain. Practical knowledge about product quality or process efficiency therefore may follow an S-shaped evolution, consistent with product or process life-cycle "interpretations." This kind of progression in product quality or manufacturing efficiency is more likely in equilibrium when the firm decides to invest a great deal on exploratory R&D. The results also carry the implication that an individual firm, or an aggregation of firms is able to maintain a steady-state in its aggregate R&D expenditures only by moving the focus of its research activities across a succession of scientific and technological domains.

By numerically analysing a specification of the model for the case of a firm developing a science-based product line developed in Chapter 3, Chapter 4 clarifies the dynamic relationships between fundamental knowledge, practical knowledge about product quality or process design and the market demand for that firm's final products. The numerical analysis rests on the use of a boundary value problem algorithm to solve the system of ordinary differential equations that are derived from the necessary conditions for a global maximum. The solution is characterised by an initially high but declining rate of investment in exploratory research, and a subsequently rising and then declining rate of application-oriented R&D. Therefore, by the end of the project planning interval the firm is completely dedicated to manufacturing with no further R&D spending.

Comparative equilibrium dynamics indicate that a firm that faces high interest rates or a small market for its product, is very inefficient in its production or research, or must pay high wages to researchers is more likely to 'cancel' its exploratory research. When these constraining conditions are very pronounced, the firm also may dispense with commercially oriented research on applications. These findings can be helpful in explaining the considerable cross-section variation in R&D performance and intensity, which is reported by empirical studies carried out at firm and line-of-business level. Numerical 'continuation

analysis' is used to investigate what happens to the optimal trajectories in the model when the firm experiences exogenous and unexpected shocks affecting market parameters or state variables during the course of a cycle of development. As the various classes of knowledge are inherited from the past, the results exhibit hysteresis effects, in that the revised R&D and production paths do not converge to the limiting levels that otherwise would have obtained had the alteration of the parameter setting occurred at the origin-point of the process.

Chapter 5 embeds the single product lifecycle model in a multiple product line context, where there are knowledge spillovers and demand-side externalities across successive product lines and fixed costs of 'switching' from one to the next. The evolution over the sequence of product lines is determined as the envelope of the optimal paths of each of them, conditional on the evolving frontier of fundamental knowledge, product quality and production technology. A solution is obtained by extending the numerical continuation methods introduced in Chapter 4. Initially, the analysis assumes that each product line is marketed for the same, exogenously given time. In this setting, the simulations reveal a growth process that traverses an ordered sequence of S-shaped product lifecycle paths. Although individual product lines run into decreasing returns to R&D investments, strong enough inter-product-line knowledge spillovers and market expansion can support a stable or increasing trend in R&D spending and a steady-state output growth path across the product lines. The results are consistent with firm-level dynamics of innovation observed by the technological forecasting literature. They can also be interpreted as having a bearing upon long-run aggregate economic growth, following Kuznets' view of the aggregate path as reflecting the successive 'flowering' of particular industries that build upon each other through demand-side or supply-side externalities.

The assumption of a predetermined time structure is relaxed later in Chapter 5 to investigate different criteria that may underlie the decision to switch between product lines.

The first extension determines the number of product lines that a profit-maximising firm should make commercially available within a fixed interval of time. The simulations reveal a trade-off in profit terms between extending the horizon of each product line, which lets the firm embark on a more ambitious R&D programme and circumvents some switching costs, and marketing several product lines, which in the presence of knowledge spillovers and growing markets may allow it to bootstrap itself onto a steeper growth path. Depending on the strength of the externalities and the fixed cost of the transitions, it may be optimal for the firm to develop and market only one product line or several. Instead, the second extension examines the potential for "accelerated product development" when this is restricted to be financially feasible. The problem is one of finding the minimum time intervals for successive cycles of product development, such that each generates sufficient profits to pay for the fixed cost of switching to the next product generation. The R&D programme meeting the test of financial feasibility has: a long planning horizon for the first product line, characterised by heavy investment in exploratory research; and much shorter lifecycles for the ensuing ones, with less R&D spending, and that mostly focused on applications-oriented research. The innovation management literature has observed this pattern of R&D spending in different science-based product families.

Chapter 6 turns to consider the interrelated challenges of modelling the irreducible element of uncertainty in the innovation process and its bearing upon the dynamics of market competition and industry structure. In an aim to capture some key aspects of Schumpeter's 'vision' of the dynamics of technologically-driven development and competition, a game theoretic approach is introduced where the structure of the industry is transformed by successive waves of innovation and imitation. The analytical framework features two methodological novelties: the explicit analysis of fundamental ('Knightian') uncertainty surrounding technological advances, in which firms' profit expectations and investment

decisions depend to some extent on a subjective evaluation of opportunities that cannot be reduced to empirical risk assessments; and the endogenous determination of market structure and R&D expenditures in a dynamic setting in which details of the sequence of entry decisions matters.

In the multistage model to which this approach leads, R&D investments in product development are made sequentially. The first stage corresponds to a mature industry, in which firms compete against each other by marketing the same "traditional" product. In the second stage, firms must decide whether or not to fund exploratory inventive research whose outcome in terms of product innovation is uncertain – it will either yield the basis for major advance in product quality, or it will yield a comparatively minor, incremental advance over existing (traditional) product designs in the industry. Other firms wait until the third stage, when the uncertainty has dissipated, before investing in costly but certain imitation of the forthcoming innovative product. These decisions are made in the context of vertical and horizontal product differentiation, where both the innovators (those who invest in the second stage) and imitators are exposed to effective competition from existing (or potential entry by) non-innovating producers of the traditional product. The decision process is represented by a subjective state-contingent pay-off matrix, in which the pay-off to alternative strategies depends upon the subsequent evolution of firm behaviours once the fundamental uncertainty of the second stage has been dispelled.

A new equilibrium concept is introduced to solve this model. The "symmetric sustainable equilibrium" is the number of symmetric firms that the market can support in each stage of the evolution, conditional on the revealed outcome of the inventive investments undertaken in the second stage. More specifically, it is the maximum number of symmetric innovators, entering imitators and traditional producers that are not forced out of the industry under each of the possible realisations of the initial inventive-innovation process. The

equilibrium is determined by combining backward recursion with zero profit and expected profit conditions for each type of firm, taking account of the leader-follower structure of sequential investments in R&D, the entry decisions by traditional producers in each period and the simultaneous quantity competition from which payoffs at each stage of the game are derived. A formal analysis is able to show that: (i) if a "symmetric sustainable equilibrium" with positive investment in stage-one innovation exists, it is unique; (ii) there can be many such "industry evolutions" – including those scenarios that display Schumpeterian aspects, such as surges (or clusters) of innovative entry followed by waves of imitation, and ensuing "creative destruction" in which traditional producers are driven out of the industry and innovators' rents are eventually eroded; (iii) not everything is possible, since exact conditions exist under which particular "historical evolutions" are infeasible; (iv) in limiting cases, small variations in the model's parameters (particularly in the fixed costs associated with innovative research, and fixed "reverse engineering" or imitation costs), can produce discrete changes ('jumps') in the equilibrium composition of firms at each stage of the evolution. An appendix extends the model to include ex post participation constraints for innovating firms.

Chapter 7 undertakes a numerical analysis of the foregoing game-theoretic model. This analysis makes use of the general uniqueness results established in Chapter 6 by treating the number of firms as continuous variables, but carry out the solutions by rounding the number of entrants in each stage to integer values; this is done by replacing the strong zero profit conditions with weak (non-negative profit) conditions throughout the solution process. The numerical results reveal a fall in competitiveness directly after innovation (in stage-two), but suggest that entry by imitating firms (in stage-three) can revert this tendency. Concentration may follow an inverted U-shaped path when innovative-R&D has produced a major innovation. To the extent that clones of the product become available, technological leaders experience a decline in market shares and there is "convergence of returns" at the

firm- and industry-level. These results seem to be consistent with empirical regularities identified in the development of new industries.

Further, the analysis suggests that major success in the innovation process appears to bring substantial benefits for consumers in two ways: first, the initial rise in consumer welfare stemming from purchases of the new, higher quality goods of the innovating firms, and second, a subsequent gain in consumer surplus when the entry of a "swarm of imitators" drives down the prices of those goods. Alternatively, a minor improvement in quality brings only marginal welfare gains to consumers because the industry becomes progressively more concentrated and prices are adjusted upwards to extract their "willingness to pay" for the incrementally better product. Other simulations indicate that under Knightian uncertainty, the firms' disposition towards uncertainty, their "optimism or "pessimism," will be important in determining industrial research intensities, and will also affect rates of entry by traditional and imitating firms.

The final chapter extends the numerical analysis of the multistage game to investigate the welfare properties of two classes of public policies: competition policies aiming to encourage greater competitiveness in the final product market, and innovation policies intending to shift the balance of R&D investments by altering the cost of new product development or imitation. The two-pronged approach by which the effects of pro-competitive policies are studied consists of (i) a comparison of the solution under Cournot (quantity) competition and Bertrand (price) competition in the product market, and (ii) comparative equilibrium dynamics with respect to the degree of horizontal product differentiation. The numerical results suggest that a policy fostering a more competitive atmosphere in the final product market will be welfare-improving, but that the added social benefits may not be all that large. This is because the industry structure endogenously becomes much more concentrated when the product market is more competitive, as firms will only finance

innovative- or imitative-R&D if they expect to recover the sunk cost of this research from their future revenue stream. The analysis holds a number of specific implications for the conduct of competition policy in industries where Schumpeterian competition – based on technological innovation – is relied upon to deliver economic welfare gains.

The fine balance struck between the equilibrium number of traditional, innovating, and imitating firms is found to be very significant to the framing of innovation policies as well. The numerical results indicate that: social welfare maximisation calls for approximately the same level of innovative-R&D as in the symmetric sustainable equilibrium; policies that stimulate imitative-R&D activities will be welfare-improving, but the increases in welfare are not so large as to justify the use of targeted subsidies to induce further entry of product clones. The implication is that government would need to foster the process of knowledge diffusion through policy instruments that do not involve direct transfers; these could take the form of information provision policies, or statutory and judicial provisions to weaken patent protection or to create exceptions for "reverse engineering." All told, the analysis implies that policy instruments need to carefully exploit the synergies between different research strategies: these may be alternative from the perspective of the firms making the investments but they are complementary from that of consumers. Chapter 8 ends by setting out a policy-relevant agenda for future work, which expressly aims to bring the dynamic modelling approaches developed in this thesis into closer conformity with the diverse 'institutional milieus' within which firms operate.

Chapter 2

Abstraction and idealization in the 'mirror of reality': A selective critique of the new growth literature

2.1 Introduction[1]

Recent work in economic growth theory, characterised as "endogenous growth" theory, has sought to break from the previous (neoclassical) tradition in macroeconomic growth models, which treated technological innovation as exogenous yet held it to be the source of sustained per capita output growth. The basic neoclassical modelling approach, originally developed by Robert Solow (1956, 1957) and Trevor Swan (1956), envisages the growth of output per capita and per worker as dependent on the process of physical (or tangible) "capital-deepening."[2] Average labour productivity levels are depicted as dependent solely on workers' access to co-operating inputs of physical capital services, and not on the level or distribution of education, experience, or other characteristics. The economy saves part of its current production and invests it in additional capital goods. Since the neoclassical model of growth assumes the existence of decreasing returns to physical capital, positive growth rates of output cannot be supported indefinitely by capital accumulation. As the capital stock grows,

[1] The material in the introduction draws on a collaborative work with Paul A. David (David with Goddard, 2001). This examined the theoretical and empirical connections between human capital acquisition and economic growth, looking in some detail at the micro-determinants of human capital acquisition, the international cross-section evidence and historical studies on modern economic growth. The reader is referred to Part II of that document for a more extensive survey of the neoclassical and new growth theories. A comprehensive list of the many survey articles in this field has been compiled and is regularly updated by Jonathan Temple. It can be accessed on the following Website: http://www.bris.ac.uk/Depts/Economics/Growth/surveys.htm.

[2] "Capital-deepening" refers usually to a rising physical capital to income ratio, whereas in some formulations, a rising ratio of capital to labour (resulting from an increasing savings to income) is referred to as increasing "capital-intensity."

the marginal productivity of investment is reduced, leading to less capacity expansion. In the limit, the economy reaches a 'steady state' in which the capital-output ratio is constant and growth in output per worker can only be maintained by the efficiency-enhancing effects of technological change. The latter source of sustained growth is thus exogenous and its productivity raising effects are treated as separable from those of capital accumulation.[3]

The endogenous growth theory emerged primarily as an attempt to encompass the sources of technological progress and therefore of sustained productivity growth within the general equilibrium framework of neoclassical theory. The principal initiator of the new trend in growth theory, Paul Romer's (1986) "AK model," generates sustained growth by assuming that productivity growth is the unintended result of the firms' investment decisions. That is, firms decide their investment without considering the positive effect that the additions in the capital stock will have on the productivity of the other firms in the economy. As a result of this externality, his model can avoid decreasing returns to physical capital from setting in. The endogenous growth models that followed Romer (1986) fleshed out the process of technological change through the explicit introduction of human capital and/or knowledge. The literature can be separated into two separate strands, depending on the role played by human capital and knowledge resources in promoting economic development.

The first line of analysis, pioneered by Lucas (1988), considers human capital to be an additional input in the production function, not fundamentally different from physical capital, although it can only be accumulated by workers through certain activities – such as education

[3] This separability feature is a consequence of the specification that technological change is "neutral," in the sense that it does not alter the *relative* marginal productivities of the factors of production, when the proportions in which labour and capital used remain unchanged. Standard growth models assume what is referred to as "Harrod neutrality," or technological progress that is purely labour-augmenting: given an unchanged capital-output ratio, such technological changes will not disturb the marginal productivity of capital inputs. The existence of this kind of technological change, and its effects in raising labour productivity, are therefore compatible with the maintenance of a "steady state" equilibrium path, along which the capital stock and output will be growing "in balance" with each other; i.e., maintaining a constant capital-output ratio and a constant real rate of return on capital.

or on-the-job training. To specify the growth path when human capital is included, it is necessary to consider an additional sector where the growth of human capital takes place. Whereas Lucas followed Ramsey (1928) in the simplifying assumption that households, as well as firms, are infinitely lived, Azariadis and Drazen (1990) and Redding (1996) undertake to model the mechanism of human capital transmission across generations in the more plausible framework of an overlapping generations model. A second line originates in the work of Nelson and Phelps (1966), and is formalised by Romer (1990) and elaborated by Young (1991), Aghion and Howitt (1992), and many others. It recognises the role of human capital as an input to conventional production but doesn't regard this as its most important function, emphasising instead the part it plays in innovation. In this perspective, human capital is the only factor of production capable of creating new and improved production processes and goods, in addition to promoting their diffusion through the economy. It is the inherent capacity to modify itself and the other inputs that can give rise to a permanently dynamic economy.

Unlike the per capita accumulation of human capital, for which there is an upper bound determined by people's finite lifespan, the new "knowledge" generated can be accumulated without bound on a per capita basis. (But note that in Becker-Barro 'dynamic models,' if human capital can be transmitted to offspring, the physical bound goes away.) There is another significant difference, in that human capital is generally taken to be rival and excludable, since it is incorporated in the workers, whereas the new "knowledge" generated is nonrival and its use is only partially excludable, to the extent that it can be made subject to intellectual property protection. A consequence of the nonrival, partially excludable nature of knowledge is that there is incomplete appropriability of the profits generated by an innovation. Even if the product or process designs cannot be used freely to produce output, they can be used to develop novel designs that erode the older goods' profits. This in turn

implies that knowledge cannot be paid its marginal product under perfect competition without negative profits for the firm.

An important contribution of Romer's (1990) model is to introduce imperfect competition into endogenous growth theory, as this generates a stream of rents to the firms when they innovate, partially circumventing this problem. In Romer's (1990) model, positive long-run growth is sustained by the accumulation of knowledge, thought of as an expansion in the varieties of intermediate capital goods. Sustained output growth is made possible by a constant returns technology in the research sector, such that increases in the "knowledge stock" do not reduce the marginal productivity of further research. Since knowledge accumulation is responsible for growth, it is the fraction of the skilled labour force in the research sector and the productivity of this pool of labour that determines the growth rate of output in the economy's steady state. In Romer's analysis, there is suboptimal knowledge accumulation and growth owing to the externality arising from the effects of knowledge on future productivity; and also from the effects of monopoly pricing in the intermediate sector, which results in suboptimal levels of acquisition of new vintage capital goods. These market failures are seen to justify government intervention. Grossman and Helpman (1991a, Ch. 3) obtain similar results by considering an expansion in the number of consumer products.

Instead, Stokey (1991, 1995), Grossman and Helpman (1991b) and Aghion and Howitt (1992) model innovation as a process that increases product quality in a predetermined number of goods. In these models, firms gain a certain probability of discovering an enhanced quality product by allocating human capital resources to R&D. This good then yields a stream of monopoly rents to the successful firm – but only until the next innovation appears. The stochastic nature of innovation means that output follows a random walk with drift such that the average growth rate is increasing in the size of the skilled workforce and in the productivity of research. As in Romer (1990), there is an inefficient

outcome because of an intertemporal knowledge spillover (as each innovation permanently raises the productivity of future goods) and a monopolistic appropriability effect. In addition, obsolescence introduces a "business stealing" effect, as firms do not internalise the destruction of rents to the current holder of a production patent when deciding their investments in R&D. Young (1993) and Stokey (1995) generalise the previous literature by showing how the equilibrium rate of R&D and hence output growth is related to the degree of substitutability or complementarity among goods and the type of competition in the industry.

To assess the strengths and weaknesses characteristic of this genre of growth theorizing, the second section of this chapter 'dissects' the 1992 model by Philippe Aghion and Peter Howitt, an elegant representative of the 'state of the art.' It is worthwhile pointing out the reasons why we have opted to review Aghion and Howitt's (henceforth A&H) theoretical framework. Apart from its mathematical elegance (or maybe, to some extent, because of it), A&H's rendition of the innovation process has become rooted in subsequent treatments elaborating this process in different directions. In effect, their dynamic framework has become a "research tool" for the analysis of innovation-driven economic growth.[4] The model's 'hip status' in and of itself would seem to warrant a re-examination of its foundations, since their strong and weak points will be inherited by the theories issuing from it. To discern these more clearly, we set A&H's model against the 'mirror of reality' afforded

[4] A measure of its influence can be ascertained by considering the substantial theoretical and empirical research that is directly based on A&H (1992). Nicholas Bloom, Richard Blundell, Rachel Griffith, Christopher Harris, Elhanan Helpman, John Vickers, and Manuel Trajtenberg are some among the many prominent economists who have extended the 'Schumpeterian growth model,' in some cases in co-authorship with A&H. See A&H (1998, 2003) for a comprehensive list of citations. Even when its exact formulation is not used, it remains a compulsory benchmark against which to compare new modelling efforts in the field. A different indicator of its sway on the corpus of knowledge dealing with technical change and economic growth has to do with the widespread inclusion of A&H's 1998 textbook, which is primarily devoted to explaining their Schumpeterian perspective, in graduate curricula in economics across the world; the other leading textbook being, of course, Barro and Sala-i-Martin (1995).

by the empirical industrial organization literature and against the framework of the "national systems of innovation" approach and the contributions in the "new economics of science."

There is a further reason why this critique is focused on A&H (1992). Their model attained its special prominence in the endogenous growth literature by its path breaking analysis of the obsolescence of knowledge. This rendered Joseph Schumpeter's idea of "creative destruction" much more precisely, and in the process brought the 'Schumpeterian system' into the fold of orthodox economic analysis – that is, an analysis with a general equilibrium framework. A natural question one would like to ask is whether (or, more exactly, in what ways) A&H's depiction of the "perennial gales of creative destruction" corresponds to Schumpeter's 'vision' of the consequences of the centrality of innovation in the growth process. With this in mind, the systematic evaluation of the constituent parts of A&H's framework is supplemented by an inquiry into the congruence of the framework taken as a whole with its purportedly Schumpeterian spirit.

The general evaluation drawn from examining the finer details of the innovation process as it is represented in the most sophisticated modellers' art casts doubt on its empirical relevance to the actual dynamics of the R&D sector. The review in Section 2.2 outlines several problematic features having to do with A&H's depiction of: (i) the innovation process and the implied evolution in knowledge resources at the economy wide level, (ii) the supply, demand, and incentives in the case of research and development personnel, and (iii) the nexus between finance and innovation. Although the 'new growth' theory had the resounding virtue of tying in the formal modelling of the R&D process with that of economic growth, it did so at the cost of biasing attention towards certain 'sources' of growth; the theory's partiality for proprietary research is one of these biases. Its faithfulness in capturing key features of Schumpeter's theory is also open to debate. A&H successfully captured the effect of innovative entry on the 'quasi-rents' generated by the proprietary

knowledge of incumbent firms. But Schumpeter specified other competitive mechanisms behind the economic obsolescence of knowledge; in particular, imitation of the technological advances was also responsible for eroding the profits of firms that invested in fundamental inventive R&D.

In part, the shortcomings exposed in each of these dimensions arise from simplifying devices permitting a formal solution by means of dynamic programming, given the restrictive assumptions needed to ensure analytic tractability with this mathematical method. But we submit that there are other, more basic limitations that are not specifically related to tractability. It is the severe and we believe unnecessary abstraction and idealization of this theory's "hard core," that is, of the set of propositions governing the purposive accumulation of valuable knowledge, which is more worrying and which this dissertation will seek to partially rectify.[5] Specifically, what this dissertation aims to deliver is a variety of solvable models with more realistic and precisely delineated microeconomic foundations as regards: the diverse objectives of the firms participating in the innovation process; the way their heterogeneous research activities are organised to meet these ends; and the fundamental uncertainty surrounding technological advances. As a necessary preliminary to this theoretical analysis, the third and last section of this chapter introduces a typology of R&D activities and knowledge resources. This typology is employed in formulating the general structure and specific assumptions of the dynamic models in Chapters 3 to 7.

[5] "Hard core" is a term introduced by Imre Lakatos in his articles on the methodology of scientific research programmes. Its meaning is succinctly described by Worrall (1998) in the *Routledge Encyclopaedia of Philosophy*: "A Lakatosian [scientific] research programme is characterized by a 'negative heuristic' principle specifying its 'hard core' (the set of basic propositions that will be implied by every theory that issues from it), and by a 'positive heuristic' (a set of directives, possibly deriving from some broad metaphysical principle, governing the construction of specific theories within the programme, and governing their modification in the light of experimental difficulties that may arise)." For the original explanation of these terms, see Lakatos (1970a, 1970b).

2.2 An examination of the 'state of the art' in new growth theory

2.2.1 An overview of Aghion and Howitt's 'creative destruction' model

The ingenious model formulated by A&H is made up of three sectors producing final goods, intermediate goods, and knowledge that is useful in making intermediate goods. Three types of workers are employed in this abstract economy: unskilled workers making final goods, skilled workers who can work in the intermediate or in the research sector, and 'specialised' workers who are exclusively concerned with R&D. The key decision in the model is the number of skilled workers employed by the firms in the research sector. This decision is taken by evaluating the expected monopoly profits from an innovation, which equal the revenue from the patent on the new, improved intermediate good minus the labour costs from skilled and specialised workers. In turn, the revenue from the patented good is equal to the discounted profits in the monopolistic intermediate sector. By employing more research workers, the probability that an innovation will emerge within a fixed interval of time increases. The innovations raise the productivity of the intermediate goods used in the production of final consumption goods. As a result, output growth is a direct function of the speed at which innovations arrive.

In the simplest specification, the number of skilled workers in research is the only variable that determines the probability that an innovation will emerge in the subsequent period. In an extension (A&H 1992, pp. 341-344), the authors allow for a second decision variable, namely, the size of the productivity jump sought by the research projects. A distinction also is made between drastic and nondrastic innovations. Drastic innovations – called "radical" innovations by Kenneth Arrow (1962), who introduced the terminology – replace the old generation of patented processes (respectively products) completely because the relative productivity (respectively utility) they afford is much greater; non-drastic innovations must still compete with previous batches of patented innovations. Notice that

although the distinction is important in the context of analyses of the consequences of the patent system, especially in terms of the incentives that the existing process or product monopolist has for further innovation, it does not correspond to the notion of "major" or "incremental" innovations; or to other distinctions based upon the innovations' macroeconomic effects on employment or investment. Since the analytical results to which these extensions lead are very similar to those obtained from the 'basic' specification, the key insights of the model can be understood from the latter.

The most significant result of the 'creative destruction' model is the determination of the equilibrium amount of research undertaken in a profit-maximising economy, which is governed by a "forward-looking difference equation, according to which the amount of research in any period depends upon the expected amount of research next period" (p. 324). This dynamic prediction hinges on two separate effects. What the authors term the 'creative destruction' effect, which reduces incentives towards current research and corresponds to the chance that a new innovation will come along in each period, thus eroding the profits of current patent-holders. And a (short-run) general equilibrium effect, also reducing the expected monopoly profits from the next patent, and leading to less research now, in this case working through the wage determination process of skilled labour. The expectation that more research will be undertaken in the following period implies a higher demand for, and hence equilibrium wages of skilled labour in that period, increasing the wage bill in the intermediate sector.

Both these negative effects work against a positive force created by the prospect of monopoly profits from new intermediate goods. It may be noticed that in essence the model leaves the strength of the positive (profit) component to be determined exogenously. There is no explicit "technology of innovation" that could lower the fixed costs of innovation. There are no tax credits or subsidies for R&D, because there is no public sector in this model.

Likewise, there are no legal changes altering the value of the monopoly rights awarded by patents. Quite naturally, with such matters left outside the basic explanatory framework, it would be tempting to correct the omission by introducing them as 'policy instruments.' In recent papers, A&H have taken that tack. A&H (1996) looks in more detail at the technology of innovation by distinguishing between "fundamental" and "secondary" innovations, and Aghion et al (2001) investigates the influence of patent policy on firms' R&D efforts.

The second important result is the characterisation of the model's stationary equilibrium, in which the allocation of skilled labour, and therefore their wage, is the same for each innovation. The authors show that in this equilibrium the log of output follows a particular stochastic process: a random walk with positive drift. The magnitude of the drift is a function of the equilibrium amount of research employment. It will be more pronounced the greater is the size of the productivity jumps, the arrival probability of each innovation, or the endowment of skilled labour; but less pronounced if the interest rate has a higher value. A&H show that other, non-stationary types of equilibrium may also exist; for example, cyclical trajectories in which the level of research oscillates between two levels in successive periods, and 'no-growth traps,' where the expectation of future research is so high that none is performed now or in subsequent periods.

There are at least two respects in which A&H's results are interesting: first, because they identify and describe analytically one source of the dynamics of R&D investments; the one working through the appropriation of monopoly profits from patented ideas. Second, because output growth is not deterministic but stochastic, in contrast to the 'old growth' literature and most of the endogenous growth models that had appeared up to 1992: in A&H, the rhythm of output growth is moved by recurring 'random bursts' of innovation.

As the A&H (1992) model is meant to capture the 'spirit' of the "knowledge-driven economy," less than the reality, the authors have left durable tangible capital out of the

picture. In principle, the intermediate goods sector could be producing capital goods. In that case, the "product innovations" for the research sector would be investment goods for the product producers; and, from the latter's standpoint, would represent "process innovations," raising the productivity in the final goods sector. But to pursue this would add considerable complications: decisions about whether to invest in the durable investments goods would be conditioned by expectations about "process obsolescence risk," and assumptions would need to be made about the competition in the "market for finance" between durable asset acquisition and R&D expenditures. Furthermore, the marginal productivity of the unskilled workers in the firms producing final goods would (presumably) be affected by the accumulation of "co-operating factors of production," and this would feed through to affect the path of real wages in the intermediate capital goods and research sectors. In short, the innovation process would be entangled with the macroeconomy as it was in a previous generation of neo-classical growth models.

What we have been given, instead, is the "pure knowledge-driven economy," in which the only capital is owned knowledge-capital, i.e., patent assets. It is a bold simplification, but, seen clearly for what it is and is not, there is little reason to expect it is immediately useful in understanding the empirical dynamics of a system in which tangible as well as intangible business assets continue to be important. The existence of yet another important intangible durable asset, human capital, which also is endogenously determined in the actual economy, constitutes another challenge to the relevance of A&H (1992) and the line of model-building that it has inspired. In A&H (1998), this imbalance has been redressed to some extent, by introducing capital accumulation into the model of creative destruction, simultaneously extending the scope of the framework to account for "a continuum of differentiated sectors." On the basis of this extension, A&H (1998, p. 114) conclude that: "[C]apital accumulation and innovation are both essential inputs to long-run growth. More innovation stimulates

capital accumulation by raising the marginal product of capital. More capital accumulation stimulates innovation by raising the profits accruing to a successful innovator."

For A&H (1992) model's results to be valid within even the narrow framework it seeks to explore, at least one and preferably both links between research investments in different periods must be thought to be significant. The 'robustness' of the creative destruction effect (as modelled by A&H) depends on the acceptance of the patent race model as a macroeconomic representation of R&D allocation decisions. It is not enough to believe in the suitability of the patent race model in the original context examined by Dasgupta and Stiglitz (1980a, b) or Dasgupta (1988a), which was circumscribed to explaining the optimal behaviour of two profit maximising firms with limited strategic possibilities, vying for a known stream of profits. (And even in this setting, questions are raised about the adequacy that the "winner takes all" in innovation.) For this genre of growth modelling to be adequate in a macroeconomic context one must believe that a patent race is a reasonable approximation of the national innovation process. This proposition is much stronger, and the evidence about the R&D process at the aggregate level, tentative though this is in some respects, does not seem to support it.

The following section aims to spell out the organisation of R&D in A&H's (1992) model in more detail and to evaluate its empirical relevance and theoretical cogency. To this end, we will compare the model's central assumptions to the characteristic features of aggregate R&D dynamics that have been identified in the empirical literature and by theorists belonging to the "national systems of innovation" approach and the "new economics of science." This comparison seems called for it one hopes to substantiate assumptions commonly used in the endogenous growth literature. We argue that the structure of the innovation process in A&H's model confronts several objections: it assumes a 'linear model' of R&D; the complementary research infrastructure is not included; and the characterisation

of knowledge is highly stylised. Each of these deficiencies will be discussed in turn. The omissions can be viewed as a legitimate reflection of the impracticalities of making a fully comprehensive formal analysis and a mark of the authors' commitment to William of Ockham's celebrated law of parsimony.[6] But we contend that a potentially tractable set of connections between the R&D process and the issues of knowledge creation and exchange and human capital formation has been suppressed. In short, the thrust of the argument is that Ockham's razor has been applied too vigorously and even imprecisely in some respects.

For the short-run, general equilibrium effect to be operative, a necessary condition is that a substantial number of "skilled workers" end up working in each of the two sectors: if the proportion of these workers in the research sector was very small, then the general equilibrium effect would be small too. A proper answer to this question calls for more knowledge about whom the "skilled workers" and "specialised workers" are meant to represent and how they relate to each other within the operation of the firms and particularly in their R&D divisions. Is one group composed of engineers and managers and the other of research workers with advanced scientific training? Or is the division of another sort? And, depending on their position and responsibility, can the rewards stipulated provide the right incentives? That in turn raises another set of questions, involving the endogeneity in the supply of research workers, as considerable costs of human capital formation must be borne to qualify them for this function. Note, by contrast, that A&H (1992) suppresses this source of dynamic complexity, making the labour supply entirely exogenous and fixed. This is why their general equilibrium effect is qualified as "short-run," only wages adjust. These questions are a matter of analysis for Section 2.2.3. How R&D is financed and who (between

[6] Summarily stated by Ockham in Latin as "pluralitas non est ponenda sine neccesitate," which directly translates as "plurality should not be assumed without necessity."

the firms, the workers, and possibly the public) bears the risk, are also critical questions for the long-run dynamics of the system, and will be briefly addressed in Section 2.2.4.

2.2.2 R&D and knowledge, in theory and in practice

Causality and feedbacks in the innovation process

A&H's schematic division of the economy into final, intermediate and research sectors is derivative of the 'supply-push' model of innovation, whereby it is the scientists who first invent something that is later taken up for development and finally commercialised. This linear model has been widely criticised because the documented experiences of innovation show more intricate patterns into which demand conditions in final product markets enter, and where the movement of people from universities to firms, and between firms, can greatly influence the transmission of knowledge and subsequent generation of commercial innovations; for citations to critiques of the 'linear model,' see David (1993b). In part, this is because A&H concentrate their attention on understanding the improvements to capital goods resulting from purposive industrial innovations. As they note in their introduction, there are "many channels through which societies accumulate knowledge, including formal education, on-the-job training, basic scientific research, learning by doing, process innovations, and product innovations" (1992, p. 323). Unfortunately, many of the models studying these channels are also based on a linear model of innovation, with no feedbacks from other stages in the research process or production, or attention to how knowledge is actually communicated from one firm to another and across sectors.

There has been a marked propensity in the economic profession to resist changes in the simple linear framework. Instead of pushing for a systemic model with feedbacks in both directions, the tendency has been to change the direction but maintain the linearity of the innovation model. Nathan Rosenberg (1974) commented on this phenomenon, arguing that

the reversal in the accepted direction of influence in the linear model of innovation during the 1960's responded in large measure to the empirical work by Jacob Schmookler. Schmookler's (1954, 1962, 1966) pioneering statistical analyses of patent data gave evidence of the strength of the connections between demand-side factors, the allocation of R&D resources, and the pace of diffusion of innovations. But, as Rosenberg discusses, Schmookler's objective was much more ambitious still: it was to demonstrate that inventive activity and the pattern of technological change were essentially economic phenomena, operating predominantly through the logic of demand-side factors.[7] Rosenberg himself was not persuaded by this more radical position, arguing instead for the coexistence of demand and supply, researchers' capabilities and the state of scientific knowledge, in any understanding of the direction and timing of inventive activities:

> "The burden of my argument here is that the allocation of inventive resources has in the past been determined jointly by demand forces which have broadly shaped the shifting payoffs to successful invention, together with supply side forces which have determined both the probability of success within any particular time frame as well as the prospective cost of producing a successful invention. But even if one were to accept the proposition, which I do not, that demand side forces alone determine the allocation of inventive resources, it would still remain true that supply side forces exercise a pervasive influence over the actual *consequences* of such resource use: i.e., the *output* of successful inventions, and the timing of these inventions."[8]

While this argument seems reasonable, it has proved difficult to corroborate. Edwin Mansfield (1991, 1995, 1998) contributed evidence in support of a 'weaker' version of this hypothesis, namely that some industrial innovations depend upon recent findings in academic research for their completion. Mansfield (1995, p. 64) came to the conclusion that a "substantial proportion of industrial innovations in high-technology industries like drugs, instruments, and information processing have been based directly on recent academic

[7] "Schmookler not only attempts to incorporate inventive activity into an economic framework. *Within* that framework he attaches overwhelming importance, as already indicated, to demand forces, and regards supply side considerations as relatively subordinate and passive." Rosenberg (1974, p. 93).
[8] Rosenberg (1974, p. 103).

research, although in many cases the invention itself did not stem from the universities." To be more precise, he estimated that upwards of 10% of the new products and processes could not have been developed without "substantial delay" lacking the knowledge generated by academic research.[9] The funding of innovation projects also revealed a "complex web of financial and intellectual relationships among academic researchers, government agencies, and firms." Although the empirical literature is still far from conclusive on the cost and time saved by innovating firms on account of supply-side factors, and a list of 'stylised facts' cannot yet be produced, the preliminary results we do have suggest that more attention should be directed to studying the potential implications, theoretical and of policy, of a systemic analysis.[10]

Public R&D and the wider infrastructure for research

If the benefits of academic research on industrial research cannot easily be measured, what is not in doubt is the magnitude of public research funding in more developed economies. To give a topical example, a report by the National Research Council (1999) estimated the U.S. federal government's research and development budget in the computer industry to have been three times as high as industrial R&D in the 1950's, falling gradually to a low of 15% of total R&D in 1979 and then rising to about 20% in the 1980's.[11] If one turns to the aggregate

[9] The strength of these linkages has been measured with other methodologies. E.g., Narin et al (1997) study the citations to scientific research papers in U.S. industry patents; 73% of the papers cited were authored at academic, governmental, and other public institutions.

[10] Metcalfe (1994) and theorists belonging to the evolutionary approach forwarded by Richard Nelson and Sidney Winter in, e.g., Nelson and Winter (1974, 1982), have broached this issue, looking in some detail at the implications of a systemic perspective for best practice in public intervention.

[11] The exact figures of the National Research Council (1999, Ch. 3) report are as follows:

> "In 1950, government funding for research and development dominated the computer world: it exceeded all industrial R&D spending on computing by a factor of three. As late as 1963, government still funded 35 percent of IBM's R&D in computing, 50 percent at Burroughs, and 40 percent at Control Data. But even by the 1960s the distribution was uneven, and several commercial suppliers, notably Honeywell and RCA, financed most of their R&D internally. Thus, the overall percentage of computer R&D supported by government declined

figures, one observes that the U.S. has amongst the highest R&D intensities at present (R&D is in the region of 2.7% of GDP), and a substantial percentage of this R&D is publicly funded (27.3%). The scale of public research funding in Japan and the EU, although smaller, is still noteworthy: in the former, R&D intensity is 2.98% and 19.6% of it corresponds to governmental spending; in the latter region, R&D intensity is 1.94%, of which 34.2% is spent by the government, although these figures conceal a significant variance among EU member countries. As an indication of the sums at stake, notice that total R&D spending in the U.S. is calculated to be 288 billion euros, so that the figure for public R&D would be approximately 79 billion euros; in the EU, the total amount spent on R&D is 164 billion euros, and therefore public R&D spending would equal about 56 billion euros.[12]

Because an important part of these resources are channelled through universities and other organisations whose ethos is remarkably different to that of firms pursuing proprietary research, the economy-wide process of innovation is actually composed of two separate but symbiotically connected 'spheres' of research. This is one of the central arguments of the "new economics of science," pioneered by Partha Dasgupta and Paul David (1987, 1994). Accordingly, special care must be taken to understand the objectives of the agents in control of these resources; for instance, military objectives were determinant in the U.S. government's decision to direct funds into the development of computer technologies. The new economics of science argues that public and proprietary research employ fundamentally different rewards and incentives. Firms typically employ monetary incentives to direct the efforts of their researchers to areas assigned high priority because of their expected private

---

dramatically from the late 1960s, both because of an absolute decline in government support and because of the rapid growth of the industry. In the mid-1970s, federal support represented only about 25 percent of computer R&D, and then shrank to a postwar low of 15 percent in 1979. With new programs and the Reagan administration's defense buildup, the level was restored to about 20 percent by 1983."

[12] The source is the DG Research of the EC, based on OECD data. It can be consulted in the European Commission Communication, "More Research for Europe," COM(2002)499, 11.09.2002.

benefits. Alternatively, the public research sector has instituted non-monetary mechanisms to motivate researchers to work in areas where high social benefits are anticipated.[13] In respect of the connections between the public and private spheres, the most obvious is that the former provides new knowledge, and transforms old knowledge so that the firms can make full use of it; but there are others, some of which flow in the opposite direction.

The omission of the 'open science system' is a major difficulty with accepting A&H's formulation of the innovation process at the national level. However, it is only the 'tip of the iceberg,' since public R&D is the most conspicuous governmental investment supporting industrial research, but the total public resources invested in the wider "research infrastructure" can be greater in volume, although harder to quantify. Apart from research funding, a country's research infrastructure is thought to include its human resources and the physical facilities wherein research can be conducted. The aforementioned National Research Council (1999) report helps us to get a sense of the share contributed by the U.S. government to each element of the research infrastructure supporting the "computer revolution."[14] In regard to human capital creation, "the percentage of graduate students in U.S. computer science and electrical engineering departments supported by federal funds (i.e., research assistantships, teaching assistantships, and fellowships) grew from 14 percent to 20 percent" between 1985 and 1996, with much higher percentages in the top computer science

[13] Most interesting among the latter is the "priority rule": by giving the 'prize' accruing from a discovery to the first person or group to have it published, the scientific community ensures that researchers and research groups are motivated to finish their work as fast as possible. The priority rule serves first to establish contests for scientific discoveries in order to go around the problem of unobservability of scientists' effort, rewarding their achievements instead. The second purpose of the mechanism is to elicit complete public disclosure of new findings. Like patents, "priority creates a privately-owned asset – a form of private property – from the very act of relinquishing exclusive possession of the new knowledge." Dasgupta and David (1987, p. 530).

[14] It is impossible to gauge the relative scale of the investments, because R&D activities and training are not accounted for separately. One reason for this is that they are by and large jointly produced. For example, researchers may produce public, codified knowledge and gain tacit knowledge simultaneously. Another example is that lecturers impart knowledge to their students whilst teaching,

departments (such as MIT, UC Berkeley, and Stanford).[15] Government spending also accounted for a sizeable part of the purchases of research equipment (65% in computer science departments between 1981 and 1995, and a similar figure in electrical engineering) and for the creation of computer facilities such as network infrastructure.

The emergence of the "national systems of innovation" approach can be interpreted as a response to the mounting body of evidence attesting to the government's key role in the innovation process, as well as that of other organisations and institutions (vis-à-vis the tendency to focus on private firms' R&D decisions in theories of technological change following in the neoclassical tradition). The contributors to this approach cast off the restraints imposed by formal modelling, and sought instead to construct heuristic frameworks that clearly delineated the goals and competences of, and the linkages between, the different actors and institutions involved in the research for innovations and their commercialisation.[16] Ultimately, they intended to examine the "structures of economic organization whose scope is far more comprehensive than the quantifiable aspects of R&D performance financed by private companies and the state, and which, through their bearing upon the generation and diffusion of innovations, contribute to the improvement of productive efficiency" (OECD, 1994, p. 9). The result of this effort has been a much better empirical understanding of: the reciprocal relations between private and public innovation; the role of finance at different stages of the innovation process; the exchange of knowledge between plants and firms in the same or different industrial clusters, regions, or countries; and, not least, science and

---

contributing to the latter's intangible human capital, but by so doing can increase their own comprehension of the subject.

[15] National Research Council (1999, Ch. 3). See this document for more details about the historical tendencies in U.S. federal funding in human resources and physical facilities and equipment. See also Nelson (1993) for comparative discussions of the role of training provided by government vs. firms; and Dasgupta and David (1994) on the allocation of personnel between university research and teaching (graduate-level) and R&D for commercial purposes.

[16] Lundvall (1992), Nelson (1993) and Edquist (1997). Roberts (1998) has written an interesting account of the 'rise' of the national systems of innovation approach.

technology policies. Theoretical models adhering to this approach have generally adopted an evolutionary set-up rather than a neoclassical one.

The final point concerning A&H's formalisation of the innovation process has to do with the assumption that each innovation has a broad, economy-wide impact. This could be justified by thinking of the innovations in question as instances of General Purpose Technologies (GPT) – such as the steam engine, the electric dynamo, and digital technology more recently. The problem is that the concept of GPT's is then reduced to its barest possible meaning, whereas its attractiveness can be thought to lie in explaining a variety of concomitant transformations in the economy. As Paul A. David has argued, a 'technological revolution' of the sort that accompanies the discovery and diffusion of a new GPT entails "a complicated techno-economic regime transition whose success is contingent upon the coordination and completion of many complementary changes in methods of production, work modes, business organization, and supporting institutional infrastructures."[17] In the light of our previous considerations, one would need to consider the adaptability of the research infrastructure existing at the time of the 'technological shock' and the government's capacity and willingness to respond to the new research and production requirements. This would address one of the items in David's characterization of the "techno-economic regime transition," but one would also need to consider the transformations undertaken by the private sector. It is the co-evolution of these sub-systems that is responsible for creating economies of scale and scope and dynamic spillovers that can eventually result in surges of productivity growth. These aspects are not captured by A&H's model of innovation-driven growth, but Helpman and Trajtenberg (1998a, 1998b) have formulated a 'Schumpeterian growth model' that tries to do so.

[17] See especially David (1989, 1990a, 1990b; 1991, 2000c).

Knowledge: A freely accessible stock?

We now turn to consider A&H's conceptualisation of knowledge resources. It seems appropriate to begin this appraisal by observing that their model differentiates between ideas exclusively by ordering them in terms of their productivity effects. This presupposes that the ideas are 'rankable' in this way ex ante. In actuality the aggregate productivity implications are uncertain, and ultimately depend upon conditions outside the sphere of control of the innovating firm: the demand conditions when the new product is finally released onto the market; the decisions of technological laggards that stand to make a profit by 'cloning' successful innovations; the legal restrictions and exceptions on the exploitation of new product designs; actions by governmental agencies to limit industrial concentration or to curb anticompetitive behaviour; etcetera. This is an instance in which 'representative firm modelling' (well-established in neo-classical tradition) continues to impose analytical blinkers on the treatment of innovation-driven macroeconomic growth by the "new growth modellers."

Still, because of the consensus that R&D is a highly risky and uncertain activity, recent growth models mediate the relationship between the 'inputs' to innovation and its 'output' by a probabilistic function. This function seeks to represent the variable effectiveness of R&D efforts in finding an innovation of known value. In A&H (1992), the random element intervening in the R&D process takes the form of a Poisson arrival rate that is a function of the economy-wide flow of research workers; this is built directly on Dasgupta and Stiglitz (1980b). In this explanation, the probability of success captures all the uncertain factors impinging on the search, discovery and development of new ideas. But the Poisson process implies that neither the amount spent on research previously nor the relative success of these investments have a bearing on current opportunities for innovation. Furthermore, A&H's treatment of risk in the search for improved intermediate goods assumes either that

knowledge is unnecessary for new designs – unlike Romer's (1990) 'AK' model, which captures the idea that a stock of knowledge, from past innovations, contributes to generating new innovations – or free for all to access in the production of new knowledge and goods. There would be complexities galore in trying to marry that formulation with the feature of "knowledge obsolescence" *à la* A&H, which presumably explains why it has been discarded. But in the process the notion of the "cumulative" nature of the process of innovation has been suppressed. Without cumulativeness, A&H proved that the log of output follows a random walk with a positive drift in a stationary equilibrium; this would not be the case if they accounted for the collective and permanent gains from prior R&D investments.[18]

But to fix this problem is not so simple. The real issue at stake is that the accumulated store of knowledge in an economy is not a simple stock and cannot be credibly represented as such. The presence of codified and tacit knowledge varieties means that a specific 'bit' of knowledge may be located in such a way as to make access to it a non-trivial matter.[19] Accessing a given piece of information may require a lengthy search; just as using it effectively may need special skills or training. If, in addition, we recognise that multiple avenues of research exist in the real world, then the relevant corpus of knowledge to a given line of research may have little or no overlap with others. Other questions can be raised about the appropriateness of the "stock of knowledge" as a modelling device. Unlike other business assets, knowledge is not a depletive resource. On the contrary, the economics of knowledge has as one of its central tenets that knowledge is not depleted when shared; see, e.g., Arrow

[18] More generally, growth models have tended to ignore learning effects derived from the search process: successful research creates knowledge that can limit or augment the possibilities for further research – even unsuccessful research can be valuable if it allows the firms to redirect its search into areas where advances may be more profitable (in the short run, at least). Likewise, nothing is said about the nature of the interaction between scientists, the importance of deliberation to the discovery of original ideas, or the process whereby discoveries are transmitted, which generally requires codification and verification by peer groups.

[19] On codified and tacit knowledge and the implications for the economic analysis of information and knowledge activities, see Cowan and Foray (1997) and Cowan, David, and Foray (2000).

(1962), Machlup (1962, 1980), and Dasgupta and David (1987, 1994). Instead, there are grounds to argument that, the more people use the "stock of knowledge," the 'better' it becomes as: any errors in the original formulation of the idea are purged; the creator's notion is restated in clearer and more articulate terms by herself or by others; and people with different backgrounds will associate a given idea with a different corpus of knowledge, and hence give it special meaning or significance.

If all of this is true, how then can the importance of an idea or invention be measured in economic terms? Many recent attempts fall back on indexes measuring the extent of its use through, e.g., citations in patents; see Trajtenberg, Henderson, and Jaffe (1997).[20] Connectedly, how can 'flows' of knowledge be identified separately from the 'stock'? Flows observed at the individual level may not constitute increases in the gross "stock of knowledge"; the latter may just be more intensively used.[21] Identifying the speed of the flows would be equally difficult, but indispensable for a model of innovation using this methodology.[22] Some institutions restrict access to an important part of the commercially useful knowledge, slowing down the flows of knowledge between people and organisations. The most conspicuous of these is the patent system. By awarding a temporary monopoly in the exploitation of original inventions, patents would increase the allocation of resources

[20] One may note in passing that knowledge also is different to natural resources (or to most commodities, for that matter) in its spatial dimension. As Samuel Johnson elegantly pointed out in *Rasselas*: "What space does the idea of a pyramid occupy more than the idea of a grain of corn?" This seems very abstract, until one considers the huge advances made in the electronic codification and compression of scientific and other types of knowledge, thus allowing for databases of previously unimaginable dimensions.

[21] A point discussed at some length in OECD (1994, pp. 33-4).

[22] Not that an accurate measurement of the speed of knowledge exchange will take us any closer to understanding if "[p]eople's ecstasy concerning the triumphs of the telegraph and the telephone often makes them overlook the fact that what really matters is the value of what one has to say, and that, compared with this, the speed or slowness of the means of communication is often a concern that could attain its present status only by usurpation"?, as Georg Simmel (1990, p. 482) wrote one century ago. Even earlier, in 1854, Henry Thoreau had expressed the same reservation: "We are in great haste to construct a magnetic telegraph from Maine to Texas; but Maine and Texas, it may be, have nothing important to communicate"; Thoreau (1999).

towards inventive activities, but also retard the process of diffusion. The optimal 'length' and 'breadth' of the patents is a matter of dispute, since there are trade-offs between providing incentives for fundamental vis-à-vis derivative innovations, and between inventive activities and those "fostering the diffusion of novel products and techniques already available for transfer."[23] The interaction between producers, "keepers," and users of various kinds of knowledge also may be restricted by firms' or other organisations' desire for secrecy.

A different oversight made by much of this literature relates to the non-monetary motivations impelling researchers to generate or exchange knowledge. Brian Loasby hypothesised that the growth of knowledge was a function of the discrepancy between our current explanatory framework and perceptual knowledge. In Loasby's (1986, p. 52) terms, it is "a response to the failure of our existing theories to predict and control, a mismatch between expectations and perceived events." This motivation has been recognised in historical studies into the emergence of systematic modes of inquiry. Geoffrey Lloyd (2002), who has investigated the "modalities of prediction" in ancient Greece and China, conjectured that the accumulation of knowledge in those societies was largely driven by the ambition to predict, explain, and ultimately control the future occurrence of different natural phenomena (lunar and solar eclipse cycles, for instance) and of the ominous events they were thought to signify (which could themselves be natural phenomena, such as droughts, but also wars, succession dates, etcetera). While these factors are non-economic in nature, it may be a mistake to leave their analysis wholly to sociologists and historians of science since an understanding of the same problem from an economic vantage point may help to frame sounder science and technology policies.

[23] David and Olsen (1992, p. 19). William Nordhaus (1967, 1969, 1972) was probably the first to undertake a mathematical analysis of patent 'length' vis-à-vis patent 'breadth.' Other key contributions to the optimal patent design literature include Klemperer (1990), Gilbert and Shapiro (1990) and Denicolo (1996).

By excluding some of the characteristic features of the aggregate innovation process, it is likely that economists' predictions about, say, the effects of industrial R&D activities on growth and welfare, will fall far from the mark. Not too long ago, an editorial in The Economist (Feb. 17 2001) boldly stated that profit and science in the much-publicised Human Genome Project enjoyed "a happy marriage," based on the private firm's (Celera) rapid sequencing of the human genome. Applying the logic in A&H, one could be misled into thinking about Celera's entry to develop and patent the techniques for sequencing, and possibly the genetic material discovered, simply in terms of a firm that created new proprietary knowledge, which then fed into higher productivity in the bio-informatics industry and so on to higher economic growth. However, such an explanation overlooks the magnitude of the public resources invested in the Human Genome Project. It was thanks to these funds that knowledge of the human genome was generated, knowledge which was indispensable to Celera when it devised its own sequencing technique. It was also this funding which provided the trained personnel that Celera employed later on. The issues at hand are not as simple as The Economist would make us believe, and the "happy marriage" between public and private research has already undergone a separation, as tensions mounted over the patenting of the genome, and seems prone to a difficult divorce.

2.2.3 The supply of and demand for research workers[24]

In A&H (1992), skilled and specialised workers can be employed in research, both contributing to an increased chance of discovering the next innovation. The Poisson arrival rate of innovations depends on a constant-returns production function that is concave in the

[24] The material in this section represents a summary of a separate paper, "Endogenous Growth Models Need an Endogenous Supply of Knowledge Workers." This is a work in progress, looking in more detail at the distinctive features of the labour market for researchers, and the collection of mechanisms aiming to influence the direction of researchers' efforts.

two types of workers and falls to zero if no skilled workers are employed. The authors do not give details of the identity of these skilled and specialised workers, or the tasks they undertake within the innovation process, but there seem to be at least two interpretations that can be put forward. First, given that the skilled workers can work in both the sector devoted to research as well as that which produces the intermediate goods, it makes sense to think of them as engineers, accountants, managers, and so on. The specialised workers who can work exclusively in the research sector are presumably scientists (physicists, chemists, mathematicians, etc.) who would find it much harder to get hired in the actual manufacturing of intermediate goods. While this interpretation is based on the assignment of the two types of workers to different occupations, a second possibility is to assign them to different 'stages' in the production of ideas. The specialised workers would be involved in the creation of knowledge that is more basic or fundamental, less adept at being commercialised, while the skilled workers assigned to the R&D division would be there to develop the 'blueprints' until they are ready to be sold.

The fact that there are two types of workers involved in research, rather than just the skilled workers, also suggests that the author's vision of research – although, once again, this is not explicitly detailed in their work – is of necessary co-operation between skilled and specialised workers: necessary because the 'research technology' is such that without 'skilled workers' the R&D sector produces no output, no matter how many specialised workers there are in the economy; co-operation since there is no discussion about how the monopoly profits from the innovation are to be divided between them (about which more anon).

A troubling feature of A&H's model is that identifying the three kinds of labour, although this adds to the realism of the model at first sight, turns out to yield little in terms of results. This is so because of the way the model is structured. First, the distribution of skills in the labour force is fixed. Second, the specialised workers, like the unskilled ones, cannot

move to another occupation or quit the labour market altogether, so that they are all employed at whatever wage optimises the system for the skilled workers, without this creating any labour supply or incentive problems at the low skill end. Third, the form of the production function implies that there are no particular ratios of specialised to skilled workers that would promote the efficiency of the economy's R&D. Fourth, workers are homogeneous in their abilities and in the effort they exert, so that the number or proportion of agents in each of the three sectors is the only significant variable. And, finally, the dividends from the patents belong to the research workers, but as these are not explicitly incorporated in the model, it is exactly equivalent in terms of the outcome for the skilled labour to be employed in either of the two sectors (intermediate goods production or research).

If the proportions of unskilled and specialised workers are fixed, and they can be employed at any wage, but only in designated sectors, then there is no real purpose to including them in the model. A choice of training (e.g., what skills to get) or employment (e.g., what sector to work in, which firm in particular, or not to work at all) would justify their inclusion; but, as noticed, human capital formation is not made endogenous. This is implicitly recognised by A&H in the book they publish in 1998: they present the model of creative destruction, except that only the skilled workers have survived from the previous version. Both the unskilled and the specialised research workers, seemingly recognised as superfluous, have been dropped out of the picture. Thus, rather than expanding the analysis to consider both human and business intangible capital, it has been restricted to just the latter. What is left after this further simplification in A&H's framework of analysis? A model that is impeccably parsimonious, in which a final sector produces consumption goods solely with intermediate goods, an intermediate sector where "leak-proof" patents are combined with skilled labour, and an R&D sector employing the same kind of skilled workers (who are sole owners of the firms) as the only input into new designs.

This portrayal has perhaps taken parsimony a cut too far: specialised researchers are crucially important to commercial innovation, for although firms need managers, accountants, production and process engineers, they are by no means sufficient. Firms cannot dispense with the need for trained personnel over the long-run, or even in the short-run. It could be argued that the prime input to innovation is the work of the specialists who are able to access and interpret the scientific and technological knowledge base, both directly through R&D in universities, government laboratories, and industry; and indirectly by teaching, thereby suppressing the need for other 'research infrastructure' inputs. These distinctions raise other questions. How are different skills formed, and what is the cost of providing the specific kinds of human capital they require?[25] What are the sources of the specialised scientific knowledge that researchers need, and is their function in the context of innovation that of translating publicly available information for its application? Or are they responsible for adding to the commonly available store of knowledge? Alternatively, are they producers of knowledge that they seek to exploit in innovation but do not share? None of these issues are broached by A&H (1992).

The relationship between human capital accumulation and technological change has since been examined by other new growth theorists, including Acemoglu (1996, 1998),

[25] The econometric studies reviewed in Stephan's (1996) survey of the scientist labour market literature (which can be referred to for specific citations) indicate that the supply of enrolees in a given field of research "almost always" depends positively on the wage in that field and negatively on the wages in other fields and occupations. This does not mean that the lags between wage signals and supply responses are not significant, or that other signals and influences in addition to wages do not influence training choices. Just as the choice of research questions can be largely non-monetary, so can the incentives that drive people to become researchers in the first place. This has been observed by sociologists such as Mulkay (1972). If the reward structure for scientists depends as much on symbolic payments through recognition of discoveries as on wages, the institutions in place (the intellectual and social 'closure' of the research communities or the conformation of 'prestige hierarchies', for instance) could be extremely important in attracting the people with the relevant capacities.

Azariadis and Drazen (1996), Eicher (1996), and Redding (1996).[26] The importance of their analyses lies in calling attention to the dynamic interactions between the two forms of investment: the potential for social increasing returns depends upon some measure of co-ordination being achieved between the investments. Importantly, this co-ordination requires an understanding of the nature of the lengthy adjustment processes governing the inflows of talented, educated individuals *and* of the demand-side, to take account of the impact of technological change on the requirements for specialised capabilities, skills and knowledge. This is because some competences needed for research may be widely substitutable across occupations and industries but other kinds are not. Such is the case with theoretical knowledge in specific subjects, whose applicability is limited to those research areas that are somehow derived or related to a given "knowledge base."[27] For that reason, the demand for researchers trained in a particular subject needs to be derived from the growth potential of the industries sharing the same knowledge base. As a corollary, it follows that technological change that increases the potential of industries requiring a particular educational background will lead to an increase in the "derived demand" for those workers – but not for other skilled workers, unless they can be retrained at a low enough cost and without too much delay.

Alternatively, when new products or processes are substitutes for older ones, the increased demand for certain types of "know-how" or "know-what" may well be accompanied by a downturn in the demand for skilled workers and researchers in other parts of the economy. This idea is at the heart of a passage written by the French novelist Albert

[26] See David with Goddard (2001) for a review of explicit treatments of the role of human capital within economic growth theory. Barro and Sala-I-Martin (1995, Ch. 5) provide an overview of two-sector models of endogenous growth.

[27] For example, physicists learn to write papers, conduct experiments, and programming, all highly general skills; but also need to become familiar with advanced mathematical methods, gravitational theory, and other subjects whose applicability is limited to research in industries requiring this specialised knowledge.

Camus, where he articulates the pains and pangs felt by those loosing their livelihood on account of the birth and growth of new industries:

> "Cooperage, threatened by the building of tankers and tank trucks, was not thriving. Fewer and fewer barrels and large casks were being made; work consisted chiefly in repairing the huge tuns already in existence. Employers saw their business compromised, to be sure, but even so they wanted to maintain a margin of profit and the easiest way still seemed to them to block wages despite the rise in living costs. What can coopers do when cooperage disappears? You don't change trades when you've gone to the trouble of learning one; this one was hard and called for a long apprenticeship."[28]

But if the social and psychological strains of technological change have been abundantly documented, confirmed facts about the changing demand for specific research skills are still lacking. Paula Stephan's (1996) survey registers some convincing evidence of a relationship between demand and R&D expenditures, and of changes in the demand for scientific personnel of particular fields brought about by disciplinary trends.

[28] The passage continues:

> "The good cooper, the one who fits his curved staves and tightens them in the fire with an iron hoop, almost hermetically, without caulking with raffia and oakum, was rare. Yvars knew this and was proud of it. Changing trades is nothing, but to give up what you know, your master craftsmanship, is not easy. A fine craft without employment and you're stuck, you have to resign yourself. But resignation isn't easy either. It was hard to have one's mouth shut, not to be able to discuss really, and to take the same road every morning with an accumulating fatigue, in order to receive at the end of every week merely what they are willing to give you, which is less and less adequate."

The short story "The Silent Men" appears in *Exile and the Kingdom*, London: Penguin, 1962. It may be tempting to discount the significance of this extract because of its emotional tenor, but the fear of long-term unemployment as technological change makes older professions obsolete is very real. A recent article in the New York Times ("As Factory Jobs Disappear, Workers Have Few Options," New York Times, September 12th 2003.) gives a remarkably similar report:

> "For nearly 30 years, Jim Greathouse was part of the blue-collar elite, a tool-and-die man who sometimes earned more than $50,000 a year because his job, making the precision metal molds that factories use to shape parts, required such exacting skills. But in June, Mr. Greathouse, 55, lost his job at the Hoover vacuum cleaner factory here, and with so many other factories laying off people, he is at a loss about what to do. "What does it do when you take away all these jobs from people who support families, who raise families?" he asked. "Manufacturing has been the strength of this country. If we can't make anything here anymore, what does that do? The fabric of this society is falling apart." "

2.2.4 Financial markets and innovation: Who wants to "fund a revolution"?

In the A&H (1992) model, profit-maximising firms decide on the number of research projects to start, and a perfect capital market supplies the investment needed for these. As A&H explain, the combination of a perfect capital market together with linear utility functions implies that the financing arrangements are immaterial; there is no need to spread the risk arising from the probabilistic innovation process because risk-neutral firms will readily bear it themselves. The authors suggest two interpretations for this capital market structure: a frictionless Walrasian credit market in which future expected consumption is discounted at a constant rate; and no credit market. In the latter case, "all nonresearch workers consume their wage income at each instant, the owners of the monopoly intermediate firm consume their flow of profit at each instant, and research workers receive no pay unless their firms innovate, at which time they are paid in shares of the next intermediate firm" (A&H 1992, p. 331).

Notice that, in equilibrium, research workers would receive an income based on the shares of the previous research firm for whom they worked. This is because the incumbent firm has no incentive to make a further investment in R&D, and no firm can acquire a segment of the market other than through patent rights. Consequently, every innovation launches a new firm and the researchers are imagined to be a floating army of specialised workers who are hired on a spot market. But, magically, although the innovations arrive randomly, the workers who receive equity in them appear not to bear any risk. How this happens remains less than fully explained. The problem is that, as David and Dasgupta (1994, p. 499) note, "scientists, like other mortals, are averse to taking risks which involve their survival and comforts." If the losers of the patent race were to get nothing, as they specify, then all the risks in the production of new knowledge would ultimately be borne by the specialised and skilled workers in the research sector.

These features of A&H's model deserve some additional comments. The distribution of returns from innovation needs to consider two elements: the wage of each kind of worker and the division of profits. As unskilled workers work exclusively in the production of final consumption goods, their wage is determined by the zero-profit condition resulting from the assumption of perfect competition in that sector. From the model, there does not seem to be any reason for unskilled workers' wage to vary, since their marginal productivity is always the same (which relates to the previous comment on the absence of tangible capital accumulation in this model). The wage paid to skilled workers follows a time path determined by a difference equation that is almost identical to the one governing the evolution of employment of these workers. The unique equilibrium wage in a stationary equilibrium can therefore be determined from the condition derived from the maximisation of expected profits for research firms and the full employment condition. Since specialised workers can only labour in the research sector, competition for jobs ensures that all are employed in equilibrium. Their wage has an upper bound determined by any financing costs and the wage bill of skilled workers. If their wage equals this notional upper bound, the owners of the patent receive zero profits; if their wage is below the upper bound, the owners obtain positive profits.

In respect to profits, the authors suggest that the workers in the R&D sector could own the "research firms," and then the "expected flow of surplus" would be divided among them. However, the ownership and control arrangements between creditors, skilled, and specialised workers are left unstated. Equally, A&H are not explicit as to the distribution of the gains from innovation. It seems from the model that increases in productivity in the intermediate goods industry translate into rising profits for the patent holders, with unskilled and specialised workers' wages remaining fixed. In the stationary equilibrium, the wages for skilled workers do not change either, further supporting the idea of increasing profits for the

shareholders of the research firms. This immediately suggests a problem of incentives if the workers in the R&D sector are sole shareholders: there would be an increasing wedge between the *income* of skilled workers in the intermediate and R&D sector, since the second would get a share of the profits in addition to their wages. The conflicts arising between salaried workers and capitalist workers were matters which classical economists took quite seriously, and which captured the attention of Piero Sraffa, Nicholas Kaldor, Joan Robinson and many other eminent economists more recently. Robinson's (1962) colourful exposition of a theory of economic growth with and without technical change is concerned precisely with the issues raised by the distribution of profits, particularly in relation to the determination of the profit and interest rate, in different "ages of growth" (golden, limping golden, leaden, galloping platinum, creeping platinum, bastard golden, and so forth). Endogenous growth theorists have only just begun to address this set of concerns.[29]

The questions raised by the division of profits between the factors of production in a firm or in the economy as a whole are serious ones, and arise even if capitals markets function perfectly well. But when capital markets are 'imperfect,' a number of additional financing and incentive problems need to be tackled. The OECD (1995) report on "National Systems for Financing Innovation" gives a good overview of the most serious concerns. The report begins by recognising that "the content of innovation-related investment and the uncertainties about its deployment create financing problems" which can be exacerbated by market failures, creating "chronic insufficiency in innovation-related investment." This "insufficiency may affect the level of investment and its contents (e.g., material versus immaterial), orientation (e.g., process versus product innovation), and distribution (e.g.,

[29] Among the exceptions, one can count Bertola (1993), who has an endogenous growth model of post-Keynesian inspiration, with a structure similar to Romer (1986). With it, he studies the "distributional implications of growth-oriented policies when agents are heterogeneous in terms of the *sources*, rather than the levels of their incomes at a point in time" (p. 1185).

SMEs versus large firms)."[30] The report singles out several factors affecting the provision of finance to firms (especially start-ups) investing in R&D. On the demand side, innovators' business skills and inclination to use the appropriate financial channels; and, on the supply side, innovators' access to the assortment of financial instruments available, knowledge of financiers who can assess their projects, and appropriate exit facilities for investors.

In an imperfect capital market, certain firm characteristics will greatly condition the need for and access to funds for research activities. Some projects for new and improved products and processes are undertaken within large companies that have long been established, and these companies can by and large finance the projects internally. Conversely, cash-constrained firms, firms that are smaller in size, and especially entrepreneurial start-ups, routinely have need of outside financing to undertake development. For the latter group, the perceived worth of their current ideas will become entangled with their past history of success or failure. Consequently, researchers' reputation could play an important role in linking a firm to appropriate finance and to other firms interested in jointly developing or manufacturing their products.

A pertinent question is then, if someone has an idea for an innovation requiring further development before it can be commercialised, what actions and complementary tangible or intangible assets are required to make this eligible for financing and, more importantly, a financially viable commercial enterprise. These two objectives may not coincide in all circumstances, since some actions or assets are better at *signalling* financial viability, while others have unobservable, yet significant effects on the innovation's commercial viability. For example, putting together a good business plan may or may not

[30] OECD (1995, p. 12). There are now numerous papers examining the influence of financial systems on industrial innovation efforts and economic growth at different levels (firm, industry, country, international cross-sections). The general conclusion of the latest survey of this empirical literature is that "the level of financial development is a good predictor of future rates of economic growth, capital

help an inventor to better understand the economic potential of his idea, but can be used as leverage to get financing from investors. The crux of the matter, as the new economics of science pointed out, is that "certain crucial inputs, such as research effort, care, innate scientific talent, and the realization of elements of chance in the process of discovery, are very costly for outsiders to monitor, and in most cases are not even observable jointly by the 'principal' who is sponsoring the inquiry (e.g. a private or public patron or employer) and the principal's 'agent' (the researcher)."[31]

The collection of mechanisms dealing with these market failures goes from formal, contractual arrangements – for example, collateral requirements or a corporate governance structure that sets out the distribution of rights and responsibilities in the enterprise – to more informal mechanisms, such as screening using past reputation or performance, or on the basis of a business plan. Thanks to the simple institutional setting assumed by A&H (1992), the representative firms can steer well clear of financing constraints. But even in affluent countries with highly developed markets for finance, many companies, and particularly recent entrants, stumble as they seek to jump over the numerous hurdles imposed by business angels, venture capitalists or banks. Because participation by financiers is essential if an R&D project is to go ahead and succeed within a reasonable amount of time, and at low enough risk to the researchers, managers, and "outside" shareholders, the division of any eventual gains and losses is a matter of the highest importance. The bottom line is that the specifics of the financing arrangements open to R&D-performing firms will have profound effects on the amount and composition of knowledge produced and on the eventual increases in productivity. Unfortunately, getting these features into macro models is something that still lies beyond the 'state of the art.'

---

accumulation, and technological change" Levine (1997, p. 689). Other surveys that can be consulted for citations to specific studies include Gertler (1988) and Pagano (1993).

2.2.5 To be or not to be Schumpeterian? What does "creative destruction" involve?

Having considered the merits and demerits of A&H's theoretical framework through an in-depth inspection of its constituent parts, this survey ends by answering an intriguing question about the legitimacy of the model's Schumpeterian provenance, an enquiry for which one must consider the framework in its entirety. Specifically, we seek to ask whether A&H's version of the episodes of "creative destruction" corresponds to Joseph Schumpeter's understanding of this process of technological change and the associated developments in employment and output at the industrial and economy wide level.[32] Alas, Schumpeter was not given an opportunity to learn about and pass judgment on the contributions that his avant-garde ideas on economic dynamics spurred, which include specific models like A&H's, but also entirely new approaches with an evolutionary perspective. Still, one may speculate about what his response would have been to A&H (1992) – providing one believes that, fifty years on, Schumpeter would still adhere to the substance of his written work, which itself is a debatable conjecture given his famous (and often criticised) inclination to rework the central building blocks of his theory during the course of his life.

In general terms, creative destruction is conceived of by A&H (2003, p. 1) as "the competitive process whereby innovative entrepreneurs introduce new technologies and products that render previous technologies obsolete." This is precised further in their model, which captures the idea that a research firm that succeeds in patenting a quality-improving innovation can force the incumbent firm from the industry, as the latter looses the monopoly

---

[31] Dasgupta and David (1994, p. 491).

[32] As an aside, one may mention that the general notion that a process of creative destruction was at work in modern capitalist economies predates Schumpeter. David A. Wells, a Harvard economist, published a book in 1889 called *Recent Economic Changes*, which stated: "[N]othing marks more clearly the rate of material progress than the rapidity with which that which is old and has been considered wealth is destroyed by the results of new inventions and discoveries." And it seems that Wells himself felt this to be no more than an economic law, previously "recognized and formulated." Michael Perelman (1995) describes the parallels between Wells' and Schumpeter's contributions.

rent on its now obsolete proprietary knowledge. A&H's characterisation comes close to Paul Strassman's (1959, p. 336) careful definition of the term:

> "With respect to technological change, it means a rate of obsolescence of unforeseen rapidity. A business is creatively destroyed if it goes bankrupt, if its capital values wither, because it has been rapidly outstripped by some other business using an innovation. Even innovating businesses may be destroyed by others with superior innovations. On a larger scale an entire industry may become unexpectedly obsolete due to innovations in other industries."[33]

For Schumpeter, however, the 'competitive gale' unleashed by innovators destroyed more than the commercial value of technological knowledge, rendering investments sunk in physical plant and equipment, business expertise and specialised craft skills no longer capable of yielding quasi-rents. As he wrote, innovation produces "an avalanche of consumers' goods that permanently deepens and widens the stream of real income, although in the first instance they spell disturbances, losses, and unemployment" (Schumpeter, 1947, p. 68). Undeniably, this insight is one of Schumpeter's most important contributions to the 'anatomy of economic change in a capitalist society.' To quote Haberler (1950, p. 365):

> "Schumpeter would not deny that some progress can be realized in a smooth and orderly fashion by continuous accumulation of capital and gradual improvement of technology. But he thinks it could not amount to much. The really big changes which shape the course of economic development are not made, cannot be made, gradually and continuously; they must be forced upon the stationary, circular flow economy in intermittent pushes."

These transitory imbalances in the economy are not addressed by A&H. The exclusive supply-side orientation of their framework and of virtually all model-building work in this area has led to a narrow focus on the characteristics of "balanced" or steady-state growth

[33] Note that Strassmann's goal in this paper was not just to clear up the meaning of the term but also to investigate its relevance in a historical episode with which Schumpeter exemplifies his theory, namely the advances made in production methods in the US between 1850 and 1914.The record of production growth from advances in power generation, ferrous metallurgy, and other industries leads him to qualify Schumpeter's argument: far-reaching novelties, although they transformed the economy, did not drive out obsolete methods, which "survived and grew for decades, though at

paths. Just as in neoclassical growth models, A&H present a simplified version of general equilibrium analysis in which changes in the quantities of outputs and inputs in the aggregate economy, and the associated relative prices, are seen to be determined simultaneously under conditions that maintain full utilisation of the economy's growing productive potential. Although the notion of steady state growth is a very useful analytical construct, in permitting the study of "comparative equilibrium dynamics," the growth theory literature is not empirically well-grounded in representing the experience of long run growth as a "steady-state equilibrium" process. Nor has it provided an adequate characterisation of the observed behaviour of the economy when it is away from its steady-state path.

The two prongs of the foregoing criticism should be understood to be quite different in their implications. The first point accepts that the long-term historical record of macroeconomic performance in the US and other advanced economies can legitimately be read as a reflection primarily of supply-side developments that gave rise to the trends in potential (full-employment) real output and productivity. But within that framework, economic historians have shown that careful reading of the evidence rejects the view of growth as a movement along an unaltered steady-state path, in favour of interpreting it as a sequence of extended transitions, or "traverses," between steady-state growth paths (see, e.g., Abramovitz and David 1973, 1999; David 1977, 2000a).

The second point of criticism focuses on the question of whether it is legitimate to separate the trend from the cycle for purposes of analysis and explanation. The neglect of the economy's *disequilibrium dynamics* – which is entailed in thus abstracting from shorter-run macroeconomic responses to exogenous demand shocks, and endogenous fluctuations in the involuntary employment rate and the rate of utilisation of fixed plant and equipment

---

reduced rates" Strassman (1959, p. 348). That is, partial obsolescence was the norm; firms with the older manufacturing systems and product designs could still find buyers for their products.

associated with the business cycle – cast doubts upon an important presupposition common among those engaged in modelling economic growth.[34] Specifically, it calls into question the validity of the claim that the steady-state growth path solution(s) found for these models are locally as well as globally stable; and that convergence to all such paths will be sufficiently rapid to provide a reasonable description of actual dynamic behaviour. If one cannot offer a good account of the economy's movements when a disequilibriating "shock" takes it away from a steady-state growth path, there may be no sound basis for assuming that the ensuing disequilibrium dynamics takes the form of simple convergence back towards the pre-existing growth path, or in a smooth traverse to a newly defined steady-state configuration.

The type of evolution that Schumpeter had in mind as an economy heads towards its "circular flow equilibrium"[35] seems entirely different: it was neither smooth, nor was it driven primarily by the actions of the group of agents that, by innovating, caused the 'technological shock' in the first place. His view was that radical innovations "are not evenly distributed in time, but that on the contrary they tend to cluster, to come about in bunches, simply because first some, and then most firms follow in the wake of successful innovation."[36] In his review of Schumpeter's *Business Cycles*, Simon Kuznets (1940, p. 259-60) put this in clearer terms: after the equilibrium is disturbed by the first entrepreneur's innovation, "his action will be followed by a swarm of imitations, increasing in volume as time passes and as the innovation becomes a more and more accepted pattern of action." In other words, Schumpeter's view is that entrepreneurial innovation periodically sets the "capitalist engine in motion," but once the engine is 'on and running,' the original innovators

[34] Cf. Fisher (1983).
[35] It is important to emphasise the function of the stationary equilibrium in Schumpeter's system. As Stolper (1951, p. 275) noticed, it "serves to isolate the moving factors and to show their proper relation to the stationary state. Static-equilibrium theory now makes sense precisely because it becomes an adjunct to a theory of development." See also Sweezy (1943).
[36] Schumpeter (1939, p. 75).

will have a limited influence over its operation. Instead, the process by which a new equilibrium is in time restored is 'powered' by the interests of other companies.

Kuznets also summarised the set of conditions, which Schumpeter called "surface factors," that intervene in the direction and speed of the 'off-the-equilibrium path.' These are "the errors of forecast; the speculative tendencies of individuals; the thousand and one peculiarities of economic institutions that are likely to prolong and exaggerate a movement once initiated." Contributors to the economics of diffusion have identified many other factors that mediate the imitation and diffusion processes. These range from purely economic factors concerning the monetary cost and time taken to 'clone' the innovation (through, e.g., reverse engineering), which are likely to be different among potential imitators; to sociological variables, such as the 'congruence' of the innovations with pre-existing practices or beliefs about their efficacy, or the "structure of social interactions" affecting access to persuasive information about the advantages of new products and production methods; to governmental policies and institutions, particularly the provision of R&D subsidies and the patent system.[37]

These considerations bring us back full circle: they suggest that A&H's formal rendition of the 'perennial gales of creative destruction' successfully captured the narrow definitions of this process given by Strassmann and others, but would not comply with broader characterisations, such as Kuznets has provided. This was the substance of Robert Solow's opinion regarding A&H's contribution:

> "Aghion and Howitt (1992) manage to give some precision to Schumpeter's vague notions about "creative destruction." They make a formal model in which each innovation kills off its predecessors. It is obvious that some innovations reduce or

[37] This partial inventory is drawn from a more inclusive evaluation of the economics of diffusion by Paul David (2003b), in a paper formalising the "micro-to-macro linkage between technological diffusion dynamics and the pace of productivity growth." This can be referred to for a list of citations to that literature.

> wipe out the rents that might otherwise have accrued to previous innovations, and this fact of life has to be taken into account in any understanding of the process."[38]

In relation to the methodological standards attained by contemporary economic theorising, Schumpeter's "notions about creative destruction" may well seem "vague." Nevertheless, they consisted of far more than the economic obsolescence of proprietary knowledge embodied in products or processes caused by the arrival of newly developed innovations.

As we mentioned already, Schumpeter also took into account the "dual nature of expansion and competition between technologies" as Bart Verspagen (2002) has put it. This is an integral part of a 'broad conception' of creative destruction and, one can argue, belongs in the "hard core" of Schumpeter's system, not in its "auxiliary belt of hypothesis" (to return to the distinctions introduced by Imre Lakatos in his methodology of scientific research programmes, on which see Footnote 5). As such, the removal of Schumpeter's disequilibrium dynamics from the process of creative destruction alters the meaning of his system in a fundamental way. Since Schumpeter's 'broad conception' is central to part of the analysis this thesis will undertake, it is convenient to quote the passage in which Verspagen (2002, pp. 2-3) explained the process:

> "There is a bandwagon of such imitations taking place during the phase after the immediate introduction of the innovation. ... The bandwagon of imitation makes the upswing happen but also implies that profit rates of the firms pushing the new technology will gradually be eroded. Initially, when there are only one or a few firms that use the new innovations, profit rates are high due to the large technological opportunities and absence of competition. But when the bandwagon grows, technological opportunities gradually become smaller (when the easiest incremental improvements have already been applied and only the harder-to-achieve improvements remain). The entrance of more and more firms on the bandwagon at the same time increases the level of competition, and this drives down the profit rate.

[38] Solow (1994, p. 52). Solow also called attention to one aspect that could be improved in A&H's model: "But sometimes—who knows, maybe just as often—innovations are complementary with predecessors and add to their rents. This possibility matters too. Is there any non-mechanical way to take both contingencies into account?" As we noted in Section 2.1, Young (1993) has developed an endogenous growth model where the equilibrium rate of R&D and hence output growth depends on the degree of substitutability or complementarity among goods.

> Eventually, this will lead to profit rates and overall growth rates settling down at the high level of the prosperity phase.
>
> Competition and the further erosion of technological opportunities do not stop in the prosperity phase, however. The continuation of these processes eventually leads to a decline in the growth rates. The recession sets in. During the recession, the technology can be considered as mature, and competition between firms mainly takes the form of (intense) price competition. The diffusion of the basic innovations gets saturated at this stage, when all potential users have adopted the technology. Hence, markets are no longer expanding and depend to a large extent on replacement of old and defective products.
>
> The recession turns into a depression when saturation gets almost complete, and the intensity of price competition reaches a peak. Technological opportunities for further improvements of the technological paradigm have dried up completely at this stage. The economy approaches a zero profit level, and the need for a new set of basic innovations becomes very high. This is when the process starts all over again with the next wave of basic innovations that may lead the economy into a new upswing."

Verspagen's account accurately summarises many of the essential points in Schumpeter's vision of creative destruction. One important aspect it overlooks concerns the 'credit mechanism.' As Elliott (1980) has highlighted in a comparative study of Schumpeter and Marx's theories of capitalist transformation, it is the supply of finance that allows "entrepreneurs to bid resources away from other uses to new innovative activities." Indeed, from Rostow's (1975, p. 720) perspective, Schumpeter manages to establish a link between price movements and technological change by "assuming that the early, experimental phase of innovation involved an inflationary expansion of credit-financed investment not matched promptly by the cost-reducing results the innovation would ultimately yield." After this episode of 'productive credit inflation,' the "maturing and diffusion of the innovation in sectors with significant weight in the economy brought cost reductions sufficiently powerful to bring down the aggregate price index."

The difficulties encountered by entrepreneurs and small firms seeking to obtain credit to fund innovative activities have already been treated in the previous section. But the point to emphasise is that, at least in his earlier writings, Schumpeter attributed the lead of the innovation process to independent entrepreneurs. For this reason, the 'normal' operation of the banking system – which at the time he was writing constituted the main source of finance

– was credited as a sine qua non condition for rapid economic change. Schumpeter's faith in the entrepreneurial endeavour and his pessimism in "Trustified Capitalism," by which he meant the larger, established firms having no shortage of funds, were subsequently criticised. The allegedly lethargic tendency of these companies was questioned shortly after the publication of *Business Cycles* by Paul Sweezy (1943, p. 96), who believed that "the form of the profit-making process itself produces the pressure to accumulate, and accumulation generates innovation as a means of preserving the profit-making mechanism and the class-structure on which it rests."[39]

In ending this inquiry we should emphasise that, whether one considers that A&H's (1992) model retained the essential elements in Schumpeter's theory of economic change or not, their call to adopt a Schumpeterian perspective has had some admirable consequences. Their appeal reinforced the movement set in motion by evolutionary theory to re-examine Schumpeter's contributions. This, in turn, encouraged the emergence of new theoretical treatments studying many of the elements that are part of a broad conception of creative destruction. One of the most interesting developments is a search-theoretic model due to Jones and Newman (1995), which aims to explain the economic cycles triggered by technological change. Instead of having a boom in output after each innovation, as in A&H (1992), the loss of sunk investments in previous technologies generates an economic downturn immediately after a fundamentally new innovation appears. Innovation renders obsolete the old technologies *and* any investments in purposive learning that created human capital specific to these technologies.[40] Alternatively, Stein (1997) solves an endogenous

[39] Later, in *Capitalism, Socialism, and Democracy*, Schumpeter changed tack, giving "big business" a major role in the process of technological change. That book's arguments stimulated a greater deal of theoretical and empirical research, reviewed in Scherer (1992).

[40] This creates a trade-off in welfare terms, as more frequent breakthroughs "promise a higher expected rate of growth in potential utility," but stability helps to avoid the one-off losses in utility from the disruptive effects of innovation on past investments. A consequence of this trade-off is that a "sufficiently long interval since the last breakthrough, or early good fortune in adapting to the current

growth model where competing firms accumulate firm-specific knowledge capital through learning-by-doing. It too leads to cycles, where there is endogenous "bunching of innovative activity." Lastly, Francois and Lloyd-Ellis (2003) extend Grossman and Helpman (1991a) to model a setting where firms can delay the implementation of their innovations. By so doing, they can explain (among other things) why 'bunching' could result from firms' profit-maximising strategies.

2.3 Stepping back from the world's complexity – a fresh approach to modelling

It has been written that in "the Navajo view of the world, language is not a mirror of reality; reality is a mirror of language."[41] Economists, however, have not been granted this freedom of expression: in the economic view of the world, reality is not a mirror of theory; theory is a mirror of reality. But it is never a perfect one. The review of A&H (1992) sought to reveal the strengths and weaknesses of its foundations – and, *mutatis mutandis*, of that genre of new growth theory – in light of the available evidence and alternative theoretical approaches. The conclusions of our inquiry, wide-ranging though it was, can be summed up in three propositions. The first, although it is almost tautological, is that new growth theory had the enormous virtue of tying in the formal modelling of the R&D process with that of economic growth; and that A&H, by admitting Schumpeterian ideas into this framework, has become one of its most inspiring variants. The second is that, whilst the specification of the labour and financial markets was highly stylised in A&H (1992), it has since been elaborated in a variety of ways (by A&H and many others). This effort has produced a huge number of

---

technology, can switch the preferred frequency of advance from maximal to minimal" (Jones and Newman 1995, p. 911).

[41] Witherspoon (1977, p. 34).

(original) testable predictions, and would certainly entitle new growth theory to be characterised as a "progressing research programme."[42]

But, and this is our third conclusion, there are obstacles to accepting A&H's (1992) representation of the innovation process itself. Seen against the 'mirror of reality,' or in contrast to other frameworks, it is clear that basic elements of the aggregate R&D dynamics have been omitted in A&H's stylised depiction. One of these is the 'open science system.' Another, the generalised assumption that R&D activities are homogenous, thus abstracting from the different classes of knowledge that can be generated by such expenditures. Both omissions are related to the reliance of this and many such models upon a "stock of knowledge," thereby conflating heterogeneous elements of knowledge having different instrumental value for different firms.[43] The severe abstraction and idealisation of these features of the innovation process is worrying because this process lies at the centre of new growth theory's "hard core." This dissertation contends that research activities and the intangible and tangible knowledge ensuing from them should be explicitly modelled. Doing so will require an appropriate dynamic framework; i.e., one that will yield a variety of solvable models with more realistic and precisely delineated microeconomic foundations. To assist in this effort, the remainder of this chapter describes a typology of privately sponsored

[42] "A research programme is said to be *progressing* as long as its theoretical growth anticipates its empirical growth, that is, as long as it keeps predicting novel facts with some success ('*progressive problemshift*'); it is *stagnating* if its theoretical growth lags behind its empirical growth, that is as long as it gives only *post-hoc* explanations either of chance discoveries or of facts anticipated by, and discovered in, a rival programme ('*degenerating problemshift*')." Lakatos (1970b, p. 100).

[43] In modern growth theory, a division is typically made between knowledge gained from purposive R&D, to which specialised resources are allocated, and 'accidental' learning processes that are a natural part of production (although recognising that certain forms of production are more conducive to a fast learning process). The difference could alternatively be viewed as between knowledge coming from a search and that originating 'spontaneously,' or in the course of other activities. Purposive or not, the knowledge gained is usually aggregated into a single 'stock,' whose measure directly determines the marginal cost of production of final goods or the utility of (and hence demand by) consumers.

R&D activities and knowledge resources that serves as a 'road map' for the development of the models in the chapters that follow.[44]

Meta-theoretical commitments

Our typology and the specific models based on it adhere to the following set of principles and concerns. We argue that privately sponsored research needs to be considered as an *uncertain but directed search for useful knowledge*. Research does not always have the same objective. R&D activities can aim to generate product and process innovations, as in the endogenous growth literature, but they can be motivated by a desire to improve a firm's research capabilities or its understanding of fundamental principles with the view of increasing the productivity of subsequent research. One also can draw a line (even if it is a fuzzy one) between a firm's research in its own products and processes, and that seeking to obtain a better understanding about others firms' innovations to develop 'clones' or to adopt cutting-edge production methods. Depending on which objective is sought, firms will have to select the appropriate mode of research, and this will leave them open to a varying degree of measurable risks and of irreducible (fundamental) uncertainties. Clearly, if one accepts that privately-backed research is heterogeneous in these aspects, the notion of an aggregate, undifferentiated stock of knowledge collapses as a modelling device. At the very least, it is necessary to keep separate track of the evolution of different stores of knowledge, according to their precise instrumental value and their role in reducing uncertainty.

The typology considers *technological progress to be cumulative in nature*. Past choices of technique, and the sunk costs associated with them exert persisting ("hysteresis")

[44] Notice that our theoretical enquiry will assume an extremely stylised institutional framework and will not examine the 'open science system.' In these respects, the critique of A&H (1992) will apply equally to the modelling effort undertaken in this dissertation. The detailed discussion in each chapter may help to dispel at least some of the reservations about the aspects excluded from the analysis.

effects upon both the rate and direction of technological improvements that are generated along established trajectories of innovation. Research tools, once acquired, can accelerate subsequent research efforts. Skills learnt by researchers as well as any disembodied knowledge permanently improve the firm's research capabilities – whether one considers the incremental improvements made in one product line or the development of different product lines, if there are strong synergies in terms of the knowledge assets required to improve them. Similarly, some components and design features become integrated to the overarching 'product concept' as time advances and the manufacturing system design is ordinarily improved in ways that condition the array of goods that it can produce. Whilst an important part of "directing" comes from the rear, one needs to acknowledge that research also responds actively to changes in demand.

Lastly, we accept the view that *knowledge is 'localised.'* Whereas intra-firm knowledge spillovers may be more or less automatic, there is persuasive evidence about the lags involved in knowledge transfers across firms; see, e.g., Mansfield (1981, 1985) and Antonelli (1995). The private and public institutions, labour market features and knowledge characteristics (especially its tacitness) impeding rapid and complete diffusion were discussed in the foregoing review. The phenomenon of localisation is examined primarily at the level of the firm, since firms will be our basic unit of analysis To be sure, knowledge can become embedded in individual plants, research teams, and researchers. Important as these lower levels of analysis are in practice, they will not form part of our theoretical framework; nor will we consider the implications of knowledge localisation for the internal organisation of the firm. However, a different form of localisation is taken into account. This is not 'anchored' in an organisation, group, or person, but rather in the product itself. Knowledge becomes localised at the product-level when there is a deliberate extension of established 'families' of products exploiting a common scientific knowledge base or a specific set of

perfected production methods. Notice that, in this case, the concept of localisation is intrinsically coupled to that of cumulativeness.

The modalities of research and knowledge – Part I

Having settled the 'operating principles' of our conceptual apparatus, we now put forward one possible arrangement of privately funded research activities and knowledge resources that, it is hoped, abides by them; the discussion of the models' structure and results in later chapters intends to supplement the description set forth here. The first division in the schema is between "exploratory research" and "applications-oriented research." Exploratory research is thought to have a derivative value to the firms, by generating "fundamental knowledge" that makes their applications-oriented research efforts more efficient, but without generating an immediate flow of revenue or a reduction in production costs. It is thought to include both disembodied knowledge and intangible knowledge that has been embodied in research platforms or tools through purposive development by the firms. Because the scope of the thesis is restricted to the R&D financed by profit-maximising firms, another appropriate label for this mode of research would be "use-inspired basic research." This category was introduced by Donald Stokes (2000) to denote research inspired by both a quest for fundamental understanding and considerations of use. He differentiates it from "pure basic research," which is concerned with the search for fundamental understanding alone, and from "pure applied research," which is primarily undertaken in consideration of its usefulness.[45]

[45] This refinement of the one-dimensional basic-applied classification is made to fit in what Stokes (2000, pp. 71-2) dubs "Pasteur's Quadrant," referring to Pasteur's dual "commitment to *understand* the microbiological processes he discovered... [and] to *control* the effects of this process on various products and on animals and humans." Our formulation also resembles, although less closely, the one prevalent in philosophy of science. For instance, Mario Bunge (1967, pp. 25-6) differentiates between two ends towards which the scientific method can be directed:

> "Primarily, to increase our knowledge (intrinsic or cognitive goal); derivatively, to increase our welfare or power (extrinsic or utilitarian goals). If a purely cognitive aim is pursued pure

Notice that the latter category is broader than our "applications-oriented research" because Stokes thinks of 'usefulness' as including more than the 'creation of value' by private firms.

Applications-oriented research is (as its name intends to emphasise) concerned with the invention and development of productive applications. The intangible knowledge embodied in the applications will generically be called "practical knowledge." In contrast to fundamental knowledge, practical knowledge is thought to have an immediate and direct influence on the firms' revenue or costs. As such, it would seem to correspond with the knowledge stock of new growth theories. But this analogy would not be completely correct, because applications are further subdivided according to whether they are "product innovations" or "process innovations." The former can go from marginal improvements in existing product designs to innovations that open entirely new product markets; however radical the improvement, they are characterised by their revenue-enhancing effects working through an enlargement in potential demand. As a simplification, we assume there is a well-defined relationship between the products' characteristics and consumers' utility. This allows the knowledge embodied in the product to be tracked by a single measure of "product quality." By contrast, R&D seeking to develop process innovations generates cost-reducing applications. The store of practical knowledge embodied in the production processes or, more generally, the 'manufacturing system design,' is summarised by one indicator alone measuring of the "efficiency of the production technology."

The parallels between this demarcation of research activities and previous typologies have been noted. But we must also acknowledge our intellectual debt to Robert Evenson and Yoav Kislev who, in their 1976 article "A Stochastic Model of Applied Research,"[46] clearly explained the dynamic complementarities between exploratory and applications-oriented

---

science is obtained. Applied science (technology) employs the same general method of pure science and several of its special methods, only applied to ends that are ultimately practical."

research within an "integrated system of production and technological research." Motivated by their studies of agricultural research and productivity, they undertook a theoretical analysis of innovation with many of the features that have been set out above; i.e., in which the value of fundamental knowledge generated by exploratory research is derived from its positive effects on the efficiency of applications-oriented, which then creates practical knowledge that is immediately advantageous to the firm's 'revenue position.' In this respect, Evenson and Kislev (hereafter E&K) anticipated the recent call made by the 'new economics of science' to follow an information-theoretic orientation when studying technological change. That is to say, to displace the focus from the inputs and outputs in innovation, to research activities and processes, and to centre attention upon the processes of generation, acquisition and distribution of knowledge.[47]

E&K's contribution goes beyond precising the affinity of these research activities. The ingenuity with which they model this relationship and the deep-seated implications of so doing are remarkable too. Instead of assuming a positive relationship between the modes of research, E&K (1976, p. 275) derive this relationship from a stochastic search process where "applied research is seen as a search within a distribution" of potential innovations with different rent streams, whereas "basic research can shift the mean of the distribution searched, it can change its variance, or it can create new technologies–discover new distributions to search." They then solve the problem through order statistics and extreme value theory. Although E&K do not seem to have noticed it, or at least they do not explicitly mention it in their work, their simple mathematical formulation managed to encompass the

---

[46] See also their book *Agricultural Research and Productivity,* published in 1975.

[47] Although they did not go into how these activities are influenced by the characteristics of knowledge (degree of codification, completeness of disclosure and ownership status) or by the norms, customs, institutions and organisations that support and mediate researchers' activities. All of the latter are also critical ingredients of the new economics of science. See Dasgupta and David (1994), and also the document prepared by Paul David and Dominique Foray for the OECD (1994), which looks at the public policy implications of this approach.

two leading ways of thinking about the effects of scientific and technical advances on the stochastic nature of the innovation process further into the future.

Firstly, the notion that R&D can reveal new opportunities for technological improvement – which may well be in previously unimagined directions, even though the research originally seemed to be run-off-the-mill and predictable in its outcomes. E&K take for granted that exploratory research is predominantly responsible for this facet of what could be called the 'research-knowledge-uncertainty triad.' So have other economists, including Richard Nelson, who stated the basis for this assumption: "While the direction of an applied research project must be closely constrained by the practical problem which must be solved, the direction of a basic research project may change markedly, opportunistically, as research proceeds and new possibilities appear."[48] Secondly, the idea that additions to the knowledge store can reduce the element of risk surrounding subsequent R&D efforts. As Kenneth Arrow (1969, pp. 29-30) candidly put it:

[48] Nelson (1959, p. 301). Nelson goes on to explain that, given the difficulty to effectively protect fundamental knowledge, firms with certain characteristics (e.g., smaller and less diversified ones) may be unable to appropriate the benefits from exploratory research and as a result may not fund this mode of research:

> "Significant advances in scientific knowledge, the types of advances that are likely to result from successful basic-research projects, very often have practical value in many fields. ... Few firms operate in so wide a field of economic activity that they are able themselves to benefit directly from all the new technological possibilities opened by the results of a successful basic research effort. In order to capture the value of the knowledge in fields which the firm is unwilling to enter, the firm must patent the practical applications and sell or lease the patents to firms in the industries affected. A firm with a narrow technological base is likely to find research profitable only at the applied end of the spectrum, where research can be directed toward solution of problems facing the firm, and where the research results can be quickly and easily translated into patentable products and processes. Such a firm is likely to capture only a small share of the social benefits created by a basic research program it sponsors. On the other hand, a firm producing a wide range of products resting on a broad technological base may well find it profitable to support research toward the basic-science end of the spectrum."

> "Technological progress is in the first instance the reduction in uncertainty. The product of a research and development effort is an observation on the world which reduces its possible range of variation. The outputs of different research projects are qualitatively different; there is no gain in acquiring the same information twice. The production of knowledge is thus basically different in character from the production of goods, where successive items can be qualitatively identical."

In E&K, as applications-oriented research finds more and more productive innovations, the likelihood of finding an even better one falls off at a rate that depends on the intensity of the search process and the properties of the distribution of potential outcomes. The ongoing search for practical knowledge is limited by the prevailing state of fundamental knowledge, as this governs the space of potential outcomes. Without continuing investment in exploratory research, decreasing returns to applications-oriented research set in, and eventually the distribution searched becomes 'technologically exhausted.' For this reason, E&K interpret this mode of research as "filling a gap between basic knowledge and the level of technology in practice."[49] Other search-theoretic models have appeared since, notably Telser (1982), Bental and Peled (1986), and Kortum (1997). These generalise E&K's results in various aspects; for instance, by introducing more general distribution functions, endogenising the production decisions, and considering multiple firms in competition with each other. But, in the process, they all drop what is arguably the most interesting feature of E&K's model, which is to highlight the unique potential afforded by basic research to avoid decreasing returns, be it in applied research or productivity growth. Without this distinction between basic and applied research modalities, it is no surprise that these other models need to introduce other elements to generate sustained output growth.[50]

[49] Evenson and Kislev (1976, p. 272). This conception of 'technological research' was originally suggested by Nelson and Phelps (1966).

[50] Which explains how they reach almost the same conclusions: "[e]ventually, the production costs are so low, continuing the research no longer pays, the research era ends, and the mature era begins" (Telser 1982, p. 69), "as the economy advances technologically it becomes harder to improve upon the state of the art technology" (Bental and Peled 1986, p. 689), "[t]echnological breakthroughs,

The dynamic models solved in Chapters 3 to 5 will take the opposite tack: they seek to capture the essential qualities of E&K's refined characterisation of the exploratory and applied modes of R&D, but the optimising problem will not be framed as a stochastic search process. Unfortunate though this methodological turnaround is, it is needed to maintain the solubility of the models in a setting where exploratory research spending and output decisions are fully endogenous. By turning to deterministic optimal control theory, the dynamic coordination of alternate R&D activities and production can be examined in more detail; and numerical methods can be introduced to study the optimised time structure of research and production in the context of multiple product lines. One should note that, without the stochastic element, the relationship between exploratory and application research studied in those chapters is closely related to Rosenberg's (1974) appreciation of these activities:

> "Thus, the growth of scientific knowledge means a gradual reduction in the cost of specific categories of science-based inventions. The timing of invention therefore needs to be understood in terms of such shifting supply curves which gradually reduce the cost of achieving certain classes of inventions."[51]

Although our theoretical examination of the dynamics of exploratory- and applications-oriented research is circumscribed to an autonomous deterministic system, Chapters 6 and 7 explore the implications of the fundamentally stochastic nature of research outcomes. To elucidate the principles guiding that theoretical inquiry, it is necessary to introduce the third partition in this typology, which turns on the diverse strategic goals that can be pursued by investment in R&D.

---

resulting in patents, become increasingly hard to find as the technological frontier advances" (Kortum 1997, p. 1389).

[51] Rosenberg's (1974, p. 107). Rosenberg also considers the interrelatedness of various kinds of fundamental knowledge and the cost of applications-R&D in different industries: "More precisely, we need to think in terms of a number of supply curves for individual industries, depending upon the knowledge bases upon which inventive activity in that industry can draw, and we need to understand more clearly the extent to which different "pools" of knowledge are potential substitutes in the inventive process" (*Ibid*). This is an interesting extension to the present classification, and to the models developed in the thesis, but one that is not explored further.

The modalities of research and knowledge – Part II

The categories of R&D activities introduced thus far sought to reflect, be it in a stylised fashion, functional differences in the knowledge they could generate, whilst simultaneously identifying the foremost complementarities between different stores of knowledge and modes of research. This classification is perhaps adequate enough if one intends to study the innovation process of an economy as a whole or that of a firm having no competitors. But if one is interested in studying the R&D spending patterns of firms in a competitive environment, one needs to distinguish between two distinct research strategies. One is motivated by firms' desire to autonomously improve their own knowledge store. The other uses R&D as a means of deciphering the knowledge previously attained by other firms or of reconfiguring it to circumvent any legal restrictions on its use. When R&D takes the form of an independent search for fundamental or practical knowledge it will be labelled as "innovative," to distinguish it from the learning processes imitating firms must finance, which are grouped under the term of "imitative-R&D" strategies. Although the latter may yield technical knowledge spillovers, their treatment in this dissertation departs from the common practice in the literature, in that they are viewed as being neither automatic nor costless, nor immediate in their effects upon the (evolving) industry structure.

It is necessary to spell out the rapport of this further partition of purposive, privately sponsored R&D activities with the other partitions introduced. We suggest that the innovative/imitative divide does not bear a one-to-one correspondence with the division of R&D into exploratory- and applications-oriented research, and the separation of the latter into product and process applications. The choice of research strategy (innovative or imitative) cannot be extricated from the choice of research activities (exploratory, product-applications, or process-applications). Supposing a 'unique profit-maximising R&D policy' could be established in a competitive scenario, its 'optimal balance' of strategic and functional

objectives would need to be determined in tandem. The outcome of this multidimensional decision process is thought to depend on structural conditions that remain more or less the same over time, and on firm characteristics that are subject to rapid change as investments in R&D reach fruition.

Among the basic structural conditions, we include: demand-side characteristics, which are typically summarised by a selection of demand elasticities (with respect to price or quality, or measuring product substitutability); the type of product market competition in the industry, which may be more or less intense depending on whether it is based on price or quality; the relative costs of capital and labour inputs, particularly of the labour services of researchers; and any sunk costs required by production or research. The characteristics of the firms at a given point in time will be synthesised by the measures of the heterogeneous stores of knowledge attained. The stores of fundamental knowledge determine the relative balance in the firms' research capabilities, whereas the two types of practical knowledge capture all the relevant product and cost characteristics. Of course, the separation between fixed structural conditions and variable firm characteristics is not as clear-cut as we have made it seem. To begin with, structural conditions are liable to change of their own accord: demand goes up and down as tastes and habits change and the economy moves through the business cycle; and so do the costs of factor inputs, in response to 'macroeconomic events' or those in their respective markets. More significantly, the prevailing structural conditions may depend on the distribution of firm characteristics in the industry; we are thinking particularly of the intensity of competition, which, as the patent race literature has shown, can become entangled with the relative standing of the firms.[52]

[52] See, for instance, the patent races modelled by Fudenberg et al (1983) or Harris and Vickers (1985, 1987). These races exhibit what has been called "Є-preemption," in which an insignificantly small lead in the innovation efforts by one firm discourages all the other firms from spending on R&D, with the consequence that the temporary leader can reinforce his position and almost surely win the race. In

The randomness associated to these research strategies is thought to be entirely different. Because innovative-R&D is a search for new fundamental or practical knowledge, it is an activity that by definition occurs beyond the established 'knowledge frontiers.' If a firm's aim is limited to making an incremental improvement to an existing product or process, the outcome will be easier to predict and 'risk management practices' may accurately assess the expected costs and benefits. But when a firm seeks to make a radical advance, in areas well beyond the established frontiers, its previous experience in research (or that of other innovating firms in the industry) will not suffice to make an objective risk assessment of the cost of reaching a given technical specification within a specified timeframe: an irreducible element of fundamental or Knightian uncertainty will persist. Quite the reverse is true of the research activities of imitating firms. By the very virtue of their purpose and scope, they will not be subject to almost any fundamental uncertainty. The goal of imitative-R&D is typically very focused: to obtain and apply the knowledge embodied in a specific research tool, product, or production technology developed through the innovative-R&D efforts of other firms. Indeed, one could define the goal of an imitative research strategy in the following way: to obtain and apply pre-existing knowledge of proven value without incurring the uncertainties in the pecuniary and opportunity costs entailed by their original development.

Locating the source of technological uncertainty in innovative research activities as against imitative ones will prove useful when we come to analyse the implications of fundamental uncertainty on the intensity and direction of industrial R&D, one of the foremost objectives of Chapters 6 and 7. The details of the game-theoretic approach with which this problem is tackled are laid out in those chapters. For the purpose of this classificatory

---

effect, the payoff functions induce intense competition for the leadership position, and once this has been established the race unwinds.

exercise, however, it is convenient to give a general indication of: the methodological difficulties of modelling fundamental uncertainty and how they will be overcome; and the implications of our approach in terms of the classification of knowledge resources that has been put forward.

Although economists have long acknowledged the ubiquitous presence of fundamental uncertainty in investment decisions, and particularly in those having to do with the generation of scientific and technical knowledge, there are few formal models of technological change incorporating this sort of uncertainty. Instead, the main stream of the literature studying technological change has restricted its analysis to risky situations in which events have objectively measurable probabilities – or has dispensed with the element of randomness altogether. This is because of a seeming lack of appropriate analytical techniques. This perception may have been fuelled by a literal interpretation of selected parts of Frank Knight and John Maynard Keynes' writings, in which they delivered their more extreme positions about uncertainty. For instance, the notion that: "About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know." But this overlooks what Keynes (1937, p. 214) wrote immediately following that statement: "Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability waiting to be summed."

The extreme notion of uncertainty fails to do justice to the subtleties of Knight and Keynes' ideas on the matter. Knight (1921, I.II.25) clearly accepted that there was a gradation between risk and uncertainty: "Some changes are foreseen and some are not, the laws of some are tolerably accurately known, of others hardly at all; and the variation in foreknowledge makes it clearly indispensable to separate its effects from those of change as

such if any real understanding of the elements of the situation is to be attained." Not only did Knight accept that a gradation existed, much of his book is devoted to examining the "structures and methods" for meeting the various types of randomness that lie in it. It is among the many insights he offers that we find a clue of how to formalise this gradation:

> The facts of life in this regard are in a superficial sense obtrusively obvious and are a matter of common observation. It is a world of change in which we live, and a world of uncertainty. We live only by knowing something about the future; while the problems of life, or of conduct at least, arise from the fact that we know so little. This is as true of business as of other spheres of activity. The essence of the situation is action according to opinion, of greater or less foundation and value, neither entire ignorance nor complete and perfect information, but partial knowledge." (Knight 1921, III.VII.5).

We owe it to Frank Ramsey, Bruno de Finetti, Leonard Savage, and to decision theorists later on to provide mathematical frameworks that could accommodate the two categories of risk and uncertainty. Among these frameworks, the most suitable one for the purpose of studying the uncertain outcomes of technological changes considers a setting of "imprecise risk"; also called a setting of "ambiguity" in the literature. This setting is such that the state of knowledge does not allow a point-estimate to be calculated, but the information can establish lower and upper bounds to the probabilities of different outcomes. Between these bounds, there remains a "grey zone" of irreducible uncertainty, as John Hicks has called it; "in it the evidence is insufficient to establish a probability" (Hicks 1979, p. 113). To determine a point-estimate, with which estimates of future profits can be made, there is no option but to introduce a subjective assessment. The firm's managers and researchers must fall back on their intuition, on their beliefs or ingrained optimism about the likelihood of success to decide the exact allocation of resources to R&D.[53]

[53] The extensive literature attempting to reconcile mathematical probability theory with subjective judgements by agents is reviewed in Shafer (1976, 1990). Within the theory of belief functions, the approach adopted in the thesis to model the agents' beliefs under uncertainty is closely related to Jean-Yves Jaffray's work on the subject.

As we will explain further in Chapter 6, modelling the stochastic nature of research in this way has an additional advantage. Since there is no requirement that the final outcome of the innovative-R&D activities coincide with the best estimate formed by the firms, or even within the bounds established *a priori*, the off-the-equilibrium paths can be examined. When the uncertainty is dispelled, the firms' subjective expectations may turn out to be wrong, leaving them with substantial losses or profits according to the discrepancy between their predictions and the actual success of their research efforts.

It remains to describe the implications of this third division of research activities (fundamentally uncertain innovative-R&D versus reasonably certain imitative-R&D) for the classification of knowledge resources. Unlike the previous demarcations, the partition into innovative- and imitative-R&D does not call for a further division of the knowledge resources in the economy. Instead, it becomes necessary to track the evolving indicators of the fundamental and practical knowledge frontiers reached by each firm through the more or less successful efforts of the R&D it has sponsored. This is because any knowledge that a given firm obtains on account of its R&D activities does not automatically spillover to the rest of the firms in the industry; as we mentioned earlier, knowledge is thought to be localised at the firm-level. This implication of knowledge localisation for model-building in the area of technological change was noticed over twenty years ago by Richard Nelson (1980, p. 65): "Drop it [the assumption of a 'common access to the blue print library'], and the analysis of productivity growth is complicated by the need to distinguish among firms in an industry and keep track of their relative weights in industry outputs and inputs." If a firm invests in innovative-R&D and succeeds in finding an improved product, the others will need to decide whether or not to fund the imitative-R&D that could secure a marketable imitation. Similarly, if a firm develops a process innovation, the then technological laggards will have to assess whether it is worth their while to undertake the costly research that would allow them to get

their hands on the knowledge embodied in the new production process, with the view of implementing it in their own plants. With that final precision, the analysis can begin.

Chapter 3

A dynamic model of exploratory- and applications-oriented research allocation by a monopoly enterprise

3.1 Introduction

Consider a firm with a product-monopoly that must decide on a research programme and a production schedule to maximise the aggregate flow of profits in a specified time frame. It has a 'product design' for a final commodity of a specific quality and a production technology, which can be exploited by employing labour in production without incurring in any research costs. But it can hire research workers to engage in either product innovation oriented research or (own)-process oriented research, both of which are modes of *applications-oriented research* N_A. Improving the product quality will raise the firm's revenue thanks to an increase in demand for the commodity it manufactures, while a superior efficiency of its process technology boosts its production of final goods for a given number of production workers. The firm can make use of the innovations that accrue from its researchers' efforts without further development or adjustment costs.

In addition to applications-oriented research, the firm can invest in *exploratory research* (N_E) to supplement the fundamental knowledge available to, and hence the productivity of research for applications. At the outset the firm's research workers already have access to some fundamental knowledge, reflecting their tacit knowledge, as well as the variety of knowledge resources that can be productively used without need of exploratory research. As with technology, once knowledge has been developed we assume that it will (costlessly) remain available for all time. Since a permanent product monopoly is assumed, it

is possible to dispense with a more sophisticated specification of the ownership rights for the "research tools" gained through exploratory research, or for the firm's product and process designs.

The complementarity between the two kinds of research implies that the decision as to how much investment to make in each should be made in tandem. Exploratory research efforts only will be profitable if enough applications-oriented research is planned in future periods, since the revenue-generating capacity of exploratory research rests on the fundamental knowledge being drawn on to create improved products or reduce production costs. *Per contra*, the productivity of applications-oriented research depends on the outcome of past expenditures to create understanding of fundamental principles, properties of materials, research tools, and "platforms," as well as skills of researchers in ways of applying these. Accordingly, the optimised trajectory of applications-oriented research must be premised on the anticipated state of fundamental knowledge at each point in time.[1] The assumption that the firm in question is a monopolist comes into play here too, as it is taken to imply that "relevant" exploratory research is not being undertaken elsewhere; or, if such knowledge is accumulating, that the process is exogenous to any action by the firm in question. This simplifying assumption is, of course, an extreme one and the implications of relaxing it will have to be examined at a later point in the analysis.

The model is set up as a constrained optimal control problem with several control and state variables. The problem is not amenable to an exact analytical solution since the resulting system of differential equations is highly nonlinear even with the simplest functional forms; nor can a qualitative analysis using phase diagrams be carried out because of the existence of

[1] The idea that current productivity (efficiency) of production depends on the history of preceding investments is articulated in Maurice Scott's 1989 book on economic growth, which held the field of "new growth theory" before Romer's contributions. Scott, however, simply related current productivity to past conventional capital formation in a sort of learning function reminiscent of Arrow's "learning by doing." The approach of this chapter is related to this, except that the capital formation history is that of the R&D activities.

several state variables. Instead, the approach followed here is to examine the profile of the dynamic paths taken by each variable assuming an interior solution exists. In this context, it is possible to establish a number of properties that must be satisfied by the optimised time paths of the firm's demand for different kinds of labour, the evolution of product quality or the efficiency of the production technology, the state of fundamental knowledge, as well as the imputed value of the state variables in profit terms. The results determine whether each of these variables is positive or negative in an interior solution and, for most of the variables, whether they would be growing or declining over time and their acceleration. This theoretical exercise helps to elucidate the differences and similarities between the research and production programmes devoted to product and process-oriented innovation.

The method and order of the chapter is the following. We begin by specifying the model's notation and assumptions in Section 3.2. Then the case of product innovation oriented research is set up in Section 3.3 as a constrained optimal control problem. The necessary and sufficient conditions for the existence of an optimal solution are determined, and we then proceed to find the profile of the optimised paths of the state, control and costate variables in an interior solution. The next section sets up the control problem for (own)-process oriented R&D and then uses the necessary and sufficient conditions to unravel the most profitable research and production programme. In Section 3.5, we make use of specific functional forms to clarify some of our earlier results and get some additional ones concerning the optimal growth rates of the different variables in the model. Finally, Section 3.6 draws out some additional implications of the analysis, reviews its limitations, and outlines the research strategy by which Chapters 4 intends to improve and extend our theoretical investigation of the dynamics of applications-oriented and exploratory R&D.

3.2 Notation and Assumptions[2]

The demand faced by the firm when it sells $q(t)$ units of a good of 'quality design' $Z(t)$ is represented by an inverse demand function $p(t)=p(Z(t),q(t))$. The inverse demand function is assumed to be continuous, differentiable as well as: (1) positive valued, $p(Z(t),q(t))>0$; (2) increasing in the product's quality $Z(t)$, $p_1(Z(t),q(t))>0$[3]; and (3) decreasing in the quantity supplied $q(t)$, $p_2(Z(t),q(t))<0$. The inverse demand is increasing in product quality because better goods give consumers greater utility, thus increasing their 'willingness to pay' at given levels of consumption.

The firm's output is determined by a production function in which manufacturing labour $L(t)$ is the only input, conditional on the production technology state $T(t)$, $q(t)=T(t)F(L(t))$. The production function $F(L(t))$ is continuous, differentiable and has the following properties: (1) labour is essential to production $F(0)=0$, (2) increasing $F'(L(t))>0$, and (3) concave $F''(L(t))\leq 0$.

Employing applications-oriented researchers $N_A(t)$ generates increases in (a) the quality of the product $\dot{Z}(t)=h(N_A(t),S(t))$, or (b) the efficiency of the production technology $\dot{T}(t)=h(N_A(t),S(t))$; these increases are conditional on the state of fundamental knowledge $S(t)$. It is true that many innovating firms conduct the two kinds of innovation simultaneously, but the analysis is simplified to firms specializing in product or manufacturing process innovations. Moreover, it is assumed that technological improvements are generated by concurrent applications-oriented research alone. This limits the applicability of the model's results, as the past history of applications-oriented research may be very significant in technological research projects. In 'sequenced' research, for instance, a number

[2] A summary of the notation employed in the model is available in the Appendix.

of components need to be developed in a particular order before the firm gains a marketable product or exploitable production technology. In those cases, the effect of research on revenue or costs would not be immediate but delayed, which would call for $h(N_A(t), S(t))$ to be a function of time or of previous investments in applications-oriented research.

To keep matters simple, the growth of product quality or process technology is represented by the same continuous, differentiable function $h(N_A(t), S(t))$, satisfying the following conditions: (1) it is nonnegative $h(N_A(t), S(t)) \geq 0$; (2) both researchers and fundamental knowledge are essential to innovation $h(0, S(t)) = h(N_A(t), 0) = 0$; (3) increasing in the two inputs $h_1(N_A(t), S(t)) > 0$ and $h_2(N_A(t), S(t)) > 0$[4]; (4) researchers and knowledge are complements in innovation $h_{12}(N_A(t), S(t)) > 0$; and, (5) jointly concave in $N_A(t)$ and $S(t)$. The last assumption is satisfied if the function is concave in the quantity of research inputs $h_{11}(N_A(t), S(t)) \leq 0$ and concave enough with respect to knowledge so that the function is jointly concave in $N_A(t)$ and $S(t)$: $h_{11}(N_A(t), S(t))h_{22}(N_A(t), S(t)) - [h_{12}(N_A(t), S(t))]^2 \geq 0$. This follows from the necessary and sufficient conditions for concavity, requiring the determinant of the Hessian of $h(N_A(t), S(t))$ to be nonnegative.[5]

Exploratory research influences technological progress indirectly, through the positive gains in the productivity of applications-oriented research resulting from new knowledge resources. The evolution of the state of fundamental knowledge available to the firm is described by $\dot{S}(t) = g(N_E(t))$, where the function $g(N_E(t))$ is twice differentiable and: (1) nonnegative $g(N_E(t)) \geq 0$, (2) research labour is essential to knowledge creation $g(0) = 0$;

---

[3] $p_i(Z(t), q(t))$ denotes the derivative of the inverse demand function with respect to argument i.

[4] $h_i(N_A(t), S(t))$ is the derivative of $h(N_A(t), S(t))$ with respect to argument i.

[5] Notice that in any interior solution the assumption of complementarity $h_{12}(N_A(t), S(t)) > 0$ will call for $h_{11}(N_A(t), S(t)) < 0$ and $h_{22}(N_A(t), S(t)) < 0$. The theorems dealing with the necessary and sufficient conditions for concavity of smooth functions and their properties are clearly stated in De la Fuente (2000, pp. 245-60).

(3) increasing $g'(N_E(t)) > 0$; and (4) concave, implying that the second derivative is less than or equal to zero $g''(N_E(t)) \leq 0$.[6]

The last set of assumptions concerns the labour markets for research workers and those employed in manufacturing. It is assumed that there are an unlimited number of workers employable in production at a constant wage w. Similarly, the supply of research workers is perfectly elastic: any number of research workers can be hired at a wage w_A if allocated to applications-oriented research, and at a wage w_E if employed to work in exploratory research. The reason for a wage differential between the two modes of research is that we intend to explore how different relative wages between these research workers influence the optimised time paths at a later stage, although one could justify it in terms of the relative scarcities of specific research skills in the labour force.

The objective of the firm is to select time paths for its controls $L(t)$, $N_A(t)$ and $N_E(t)$ to maximise the sum of profits from the initial time t_0 to the terminal time t_1,

$$\Pi = \int_{t_0}^{t_1} p(Z(t), q(t))q(t) - wL(t) - w_A N_A(t) - w_E N_E(t)dt$$

where $q(t) = T(t)F(L(t))$. This maximisation is subject to:

1) either $\dot{Z} = h(N_A(t), S(t))$ or $\dot{T} = h(N_A(t), S(t))$, depending on whether the firm is orienting its research towards improving product quality or manufacturing efficiency,

[6] The sufficient conditions for the existence of a global maximum to the dynamic problem require the 'production functions' for fundamental knowledge and the 'practical knowledge' embodied in the improved products and processes to be concave, but it is not obvious that this is the prevailing state facing innovating firms in practice. One can account for the existence of concavity – for instance, the quality of researchers' efforts may decline if they are pressed to work longer hours, or they may reach impasses when solving more ambitious problems; but also for convexity, as research workers may be undergoing a learning process that improves their capabilities, or communication flows between researchers could improve as a research project becomes more advanced.

2) $\dot{S} = g(N_E(t))$, the equation of motion describing the relationship between the state of fundamental knowledge and exploratory research inputs,

3) nonnegativity constraints on the three control variables $L(t) \geq 0$, $N_A(t) \geq 0$ and $N_E(t) \geq 0$,

4) and boundary conditions on the state variables, namely the initial value of product quality $Z(t_0) = Z_0 > 0$ and the state of the production technology $T(t_0) = T_0 > 0$, as well as that of the state of knowledge $S(t_0) = S_0 > 0$, leaving the terminal values $Z(t_1)$, $T(t_1)$ and $S(t_1)$ free.

3.3 Product quality innovation

The problem just described is a constrained control problem with a finite planning horizon and free endpoints for the state variables, consisting in finding the optimal path $L^*(t)$, $N_A^*(t)$ and $N_E^*(t)$ for the control variables so as to maximise profits. To simplify the notation, we assume that the state of the process technology is $T(t_0) = T(t) = 1$ throughout this section. The Hamiltonian function for the problem is:

$$\begin{aligned} H &= H(Z(t), S(t), L(t), N_A(t), N_E(t), \pi_Z(t), \pi_S(t)) \\ &= p(Z(t), F(L(t)))F(L(t)) - wL(t) - w_A N_A(t) - w_E N_E(t) \\ &\quad + \pi_Z(t)h(N_A(t), S(t)) + \pi_S(t)g(N_E(t)) \end{aligned} \tag{3.1}$$

where $\pi_Z(t)$ is the costate variable corresponding to the differential equation of motion $\dot{Z}(t)$, a measure of the imputed value of an increase in product quality in profit terms. Analogously, $\pi_S(t)$ is the costate variable associated to $\dot{S}(t)$, interpreted as the imputed value of an increase in knowledge resources in profit terms. One can then define a Lagrangian function,

$$\mathsf{L} = H + \lambda_L(t)L(t) + \lambda_{N_A}(t)N_A(t) + \lambda_{N_E}(t)N_E(t)$$

$$= p(Z(t),F(L(t)))F(L(t)) - wL(t) - w_A N_A(t) - w_E N_E(t) + \pi_Z(t)h(N_A(t),S(t)) \qquad (3.2)$$
$$+ \pi_S(t)g(N_E(t)) + \lambda_L(t)L(t) + \lambda_{N_A}(t)N_A(t) + \lambda_{N_E}(t)N_E(t)$$

where $\lambda_L(t)$, $\lambda_{N_A}(t)$ and $\lambda_{N_E}(t)$ are Lagrange multipliers corresponding to the three nonnegativity constraints on $L(t)$, $N_A(t)$ and $N_E(t)$ respectively. The optimal solution must satisfy the following necessary conditions (so as to simplify the notation we get rid of the (t)'s and suppress the asterisk for the control, state and costate variables, but one needs to bear in mind that they refer to the optimal paths):[7]

$$\frac{\partial L^*}{\partial L} = p_2(Z,q)F'(L)q + p(Z,q)F'(L) - w + \lambda_L = 0 \qquad (3.3)$$

$$\frac{\partial L^*}{\partial N_A} = -w_A + \pi_Z h_1(N_A,S) + \lambda_{N_A} = 0 \qquad (3.4)$$

$$\frac{\partial L^*}{\partial N_E} = -w_E + \pi_S g'(N_E) + \lambda_{N_E} = 0 \qquad (3.5)$$

$$\lambda_L \geq 0 \qquad L \geq 0 \qquad \lambda_L L = 0$$
$$\lambda_{N_A} \geq 0 \qquad N_A \geq 0 \qquad \lambda_{N_A} N_A = 0 \qquad (3.6)$$
$$\lambda_{N_E} \geq 0 \qquad N_E \geq 0 \qquad \lambda_{N_E} N_E = 0$$

$$\dot{\pi}_Z = -\frac{\partial L^*}{\partial Z} = -p_1(Z,q)q \qquad (3.7)$$

$$\dot{\pi}_S = -\frac{\partial L^*}{\partial S} = -\pi_Z h_2(N_A,S) \qquad (3.8)$$

[7] The necessity and sufficiency conditions obtained in this and the following section follow the necessity and sufficiency theorems for a general constrained problem that are explained by Léonard and Van Long (1992, pp. 210-18). A more intuitive formulation of these conditions has been made by Chiang (1992) and, on the other side of the spectrum, there are rigorous proofs in, e.g., Seierstad and Sydsaeter (1977) or Hart et al (1995).

$$\dot{Z} = \frac{\partial \mathsf{L}^*}{\partial \pi_Z} = h(N_A, S) \tag{3.9}$$

$$\dot{S} = \frac{\partial \mathsf{L}^*}{\partial \pi_S} = g(N_E) \tag{3.10}$$

$$Z(t_0) = Z_0 > 0 \qquad \pi_Z(t_1) = 0 \tag{3.11}$$

$$S(t_0) = S_0 > 0 \qquad \pi_S(t_1) = 0 \tag{3.12}$$

The complementary slackness conditions in (3.6) require either the control variables or the associated Lagrange multipliers to equal zero at every point in time. If the nonnegativity constraints are not binding, the Lagrange multipliers take a value of zero and one can put forward more intuitive interpretations about the conditions with respect to each control variable. In that case, the first order condition with respect to production labour in equation (3.3) gives the correspondence between the employment of manufacturing workers, their wage and product quality that must hold at each point in time; this is such that the marginal revenue product of labour equals its marginal cost, the wage w. The marginal revenue product of labour has two components: the gains accruing to the firm from the additional output sold in the market (second term in equation 3.3); and the loss in revenue resulting from a reduction in the willingness to pay as supply increases (first term in 3.3).

The first order condition with respect to applications research inputs, in equation (3.4), indicates that the intensity of the research effort developing new product designs should be such that its marginal benefit – the marginal quality increase times the imputed value of this improvement – equals the marginal cost, the wage w_A. The same is true of condition (3.5) for exploratory research: its level is optimised when the wage of exploratory researchers w_E equals the valuation of the increase in knowledge from the extra researchers.

Equation (3.7) shows that the value imputed to product quality innovations by the firm would fall over time at a rate that exactly equals the marginal revenue product with respect to product quality:

$$\dot{\pi}_Z = -p_1(Z,q)q = -\frac{\partial[p(Z,q)q]}{\partial Z}$$

The rate of change in the shadow price of additions to the fundamental knowledge base (equation 3.8) is tied to the imputed value of product quality multiplied by the marginal quality increase resulting from additional knowledge $\pi_Z h_2(N_A,S)$. This condition is obtained because, as Chapter 2 and the introduction of this chapter explained, this class of knowledge resources has a derivative or subordinate economic value to the firm. To the extent that practical knowledge embodied in applications is valuable, additions to the fundamental knowledge store will be so too; but the optimal intensity of the exploratory research efforts also needs to factor in the productivity of research workers in this activity.

Finally, $\pi_Z(t_1)=0$ and $\pi_S(t_1)=0$ (in 3.11 and 3.12, respectively) are the necessary transversality conditions for a problem with free-endpoints for the state variables.

Notice that the state of knowledge and product quality will be positive and at least as great as the initial values at any point in time for any admissible control trajectory (one where N_A and N_E are nonnegative). This follows immediately from the assumption that the states of knowledge and product quality are positive at t_0, $S(t_0)=S_0>0$ and $Z(t_0)=Z_0>0$, and the properties of $g(N_E)$ and $h(N_A,S)$.[8]

[8] $N_E \geq 0$, $g(0)=0$, $g'(N_E)>0$ and $\dot{S}=g(N_E) \Rightarrow \dot{S}\geq 0$
$\dot{S}\geq 0$ and $S(t_0)=S_0>0 \Rightarrow S\geq S_0>0$
$N_A \geq 0$, $S>0$, $h(0,S)=0$, $h_1(N_A,S)>0$ and $\dot{Z}=h(N_A,S) \Rightarrow \dot{Z}\geq 0$,
$\dot{Z}\geq 0$ and $Z(t_0)=Z_0>0 \Rightarrow Z\geq Z_0>0$

If the Lagrangian is concave in the control and state variables, the necessary conditions listed above are sufficient for a global maximum of the problem. Léonard and Van Long (1992, p. 214) show that the Lagrangian will be concave if (i) the objective functional, (ii) each of the equations of motion and (iii) each of the inequality constraints is jointly concave in the control and state variables. Condition (ii) is satisfied from the assumptions of concavity of $h(N_A, S)$ and $g(N_E)$, and (iii) is met because the nonnegativity constraints are linear and hence concave.

We are left with condition (i), requiring the objective functional, which in the problem is $p(Z, F(L))F(L) - wL - w_A N_A - w_E N_E$, to be jointly concave in Z, S, L, N_A and N_E. It does not depend on S and is related linearly to and hence is concave in N_A and N_E. Taking note that the sum of concave functions is concave, it remains to impose conditions on the inverse demand and the production function so that revenue $p(Z, F(L))F(L)$ is jointly concave in Z and L. We will assume that revenue is strictly concave to get clearer cut results, but it should be clear from the analysis that the majority of the derivations do not depend on making this slightly stronger assumption. A sufficient condition for revenue to be strictly concave is that the Hessian of $p(Z, F(L))F(L)$ be negative definite for any Z and L in the relevant domain. Negative definiteness is guaranteed if the second derivative of revenue with respect to product quality is less than zero,

$$p_{11}(Z,q)q < 0$$

and the determinant of the Hessian matrix is greater than zero,

$$p_{11}(Z,q)q\left\langle p_2(Z,q)qF''(L) + p(Z,q)F''(L) + p_{22}(Z,q)qF'(L)^2 + 2p_2(Z,q)F'(L)^2 \right\rangle$$

$$-\left[p_{12}(Z,q)qF'(L) + p_1(Z,q)F'(L)\right]^2 > 0$$

or, rearranging,

$$F''(L)[p_2(Z,q)q + p(Z,q)] + F'(L)^2[p_{22}(Z,q)q + 2p_2(Z,q)] < \frac{F'(L)^2[p_{12}(Z,q)q + p_1(Z,q)]^2}{p_{11}(Z,q)q} \leq 0$$

The two conditions restrict the shape of the inverse demand function and the production function. It is not apparent what those limitations are in this general setting, but they can be readily understood once specific functional forms are introduced in Section 3.5 and we postpone an interpretation until that point.

Although the existence of a solution is guaranteed by the above considerations, the existence of an *interior solution* can only be ensured by imposing an equivalent of the Inada conditions in the Solow-Swan growth model; in this case for $h(N_A, S)$ and $g(N_E)$ as well as for the production function $F(L)$. But the assumption that the marginal productivity of the two research activities tends to infinity as the number of research workers employed drops to zero, or that it would tend to zero as R&D inputs increase towards infinity, cannot be justified empirically at the firm and line-of-business level.[9] Instead of making rather implausible assumptions about researchers' productivity to ensure an interior solution, we follow other authors employing optimal control theory in the context of innovation in assuming that (given the constellation of wages, initial conditions, etcetera) an interior solution exists in the relevant planning horizon and characterizing the solution path in that interval; see, e.g., Kamien and Schwartz (1972a, b). We should point out that Chapter 4 undertakes a numerical comparative dynamics analysis of a specified version of this model that provides many

[9] In the real world, many firms do not invest in applications-oriented R&D, let alone exploratory R&D – they decide or, one might argue, are forced by their environment to select a corner solution. This is a robust finding emerging from innovation surveys across industries and countries. Even in countries where the ratio of R&D spending to GDP is greatest, there is a significant percentage of firms that do not fund any R&D. See, e.g., Cohen and Klepper (1992), who found that about 16% of firms in the U.S. manufacturing sector reported making no expenditure on R&D. And in their study it is clear that R&D intensities vary a great deal from one industry to another: the percentage of firms with zero R&D intensity ranged from 5% in machinery to over 50% in printing and publishing.

insights about the conditions under which one may expect the firm to select a corner solution over an interior one.

Suppose, then, that the three control variables are positive in the relevant planning horizon, $L>0$, $N_A>0$ and $N_E>0$. In that case, it is possible to establish a number of properties that must be satisfied by the optimised time paths of the firm's demand for different kinds of labour, the evolution of the product quality and the state of knowledge, as well as the imputed value of the two state variables in profit terms. In the interest of clarity and completeness, we first verify some simple results following directly from the assumption of an interior solution, and then go on to those requiring more elaborate derivations. The specific assumptions and mathematical operations needed to get each result are stated in brief terms.

(R1) The Lagrange multipliers are identically equal to zero $\lambda_L=0$, $\lambda_{N_A}=0$, $\lambda_{N_E}=0$

$L>0$ and $\lambda_L L=0 \Rightarrow \lambda_L=0$

$N_A>0$ and $\lambda_{N_A}N_A=0 \Rightarrow \lambda_{N_A}=0$

$N_E>0$ and $\lambda_{N_E}N_E=0 \Rightarrow \lambda_{N_E}=0$

(R2) The firm's output is positive $q>0$

$L>0$, $q=F(L)$, $F(0)=0$ and $F'(L)>0 \Rightarrow q>0$

(R3) The price set is positive $p=p(Z,q)>0$

$\lambda_L=0$ and $p_2(Z,q)F'(L)q+p(Z,q)F'(L)-w+\lambda_L=0 \Rightarrow p(Z,q)=\dfrac{w}{F'(L)}-p_2(Z,q)q$

$w>0$, $F'(L)>0$, $p_2(Z,q)<0$ and $q>0 \Rightarrow p(Z,q)=\dfrac{w}{F'(L)}-p_2(Z,q)q>0$

(R4) The shadow price of additions to the knowledge base is positive $\pi_S > 0$

$\lambda_{N_E} = 0$ and $-w_E + \pi_S g'(N_E) + \lambda_{N_E} = 0 \Rightarrow \pi_S = \dfrac{w_E}{g'(N_E)}$

$w_E > 0$ and $g'(N_E) > 0 \Rightarrow \pi_S = \dfrac{w_E}{g'(N_E)} > 0$

(R5) The shadow price of product quality improvements is positive $\pi_Z > 0$ [10]

$\lambda_{N_A} = 0$ and $-w_A + \pi_Z h_1(N_A, S) + \lambda_{N_A} = 0 \Rightarrow \pi_Z = \dfrac{w_A}{h_1(N_A, S)}$

$w_A > 0$ and $h_1(N_A, S) > 0 \Rightarrow \pi_Z = \dfrac{w_A}{h_1(N_A, S)} > 0$

(R6) Fundamental knowledge resources are growing over time $\dot{S} > 0$

$N_E > 0$, $\dot{S} = g(N_E)$ with $g(0) = 0$ and $g'(N_E) > 0 \Rightarrow \dot{S} > 0$

(R7) The quality of the good sold is increasing over time $\dot{Z} > 0$

$N_A > 0$, $S > 0$ and $\dot{Z} = h(N_A, S)$ with $h(0, S) = 0$ and $h_1(N_A, S) > 0 \Rightarrow \dot{Z} > 0$

(R8) A constant or decreasing output $\dot{q} \leq 0$ is sufficient for the price to be increasing over time $\dot{p} > 0$; $\dot{q} > 0$ is necessary but not sufficient for $\dot{p} \leq 0$; $\dot{p} \leq 0$ is sufficient for $\dot{q} > 0$

Differentiating $p = p(Z, q)$ with respect to time gives $\dot{p} = p_1(Z, q)\dot{Z} + p_2(Z, q)\dot{q}$

$p_1(Z, q) > 0$, $\dot{Z} > 0$, $p_2(Z, q) < 0$ and $\dot{p} = p_1(Z, q)\dot{Z} + p_2(Z, q)\dot{q} \Rightarrow \dot{q} \leq 0$ is sufficient for $\dot{p} > 0$, $\dot{q} > 0$ is necessary for $\dot{p} \leq 0$ and $\dot{p} \leq 0$ is sufficient for $\dot{q} > 0$

[10] Results (4) and (5) are essential, in that the Lagrangian will be concave if the differential functions are concave, *provided* the optimal value of the costate variables is nonnegative.

(R9) The rate of change of output has the same sign as the rate of change of employment in manufacturing $Sign[\dot{q}] = Sign[\dot{L}]$

Differentiating $q = F(L)$ with respect to time gives: $\dot{q} = F'(L)\dot{L}$

$F'(L) > 0$ and $\dot{q} = F'(L)\dot{L} \Rightarrow Sign[\dot{q}] = Sign[\dot{L}]$

(R10) A constant or decreasing employment in manufacturing $\dot{L} \leq 0$ is sufficient for the price to be increasing over time $\dot{p} > 0$; $\dot{L} > 0$ is necessary but not sufficient for $\dot{p} \leq 0$; $\dot{p} \leq 0$ is sufficient for $\dot{L} > 0$

Differentiating $p = p(Z, F(L))$ with respect to time gives $\dot{p} = p_1(Z,q)\dot{Z} + p_2(Z,q)F'(L)\dot{L}$

$p_1(Z,q) > 0$, $p_2(Z,q) < 0$, $\dot{Z} > 0$, $F'(L) > 0$ and $\dot{p} = p_1(Z,q)\dot{Z} + p_2(Z,q)F'(L)\dot{L} \Rightarrow \dot{L} \leq 0$ is sufficient for $\dot{p} > 0$, $\dot{L} > 0$ is necessary for $\dot{p} \leq 0$ and $\dot{p} \leq 0$ is sufficient for $\dot{L} > 0$

(R11) The shadow price of additions to the knowledge base decreases over time $\dot{\pi}_S < 0$

$\pi_Z > 0$, $h_2(N_A, S) > 0$ and $\dot{\pi}_S = -\pi_Z h_2(N_A, S) \Rightarrow \dot{\pi}_S < 0$

(R12) The shadow price of quality improvements decreases over time $\dot{\pi}_Z < 0$

$q > 0$, $p_1(Z,q) > 0$ and $\dot{\pi}_Z = -p_1(Z,q)q \Rightarrow \dot{\pi}_Z < 0$

(R13) The rate of change of employment in manufacturing has the same sign as $p_{12}(Z,q)q + p_1(Z,q)$; $p_{12}(Z,q) \geq 0$ is sufficient for $\dot{L} > 0$; $p_{12}(Z,q) < 0$ is necessary but not sufficient for $\dot{L} \leq 0$

$\lambda_L = 0$ and $p_2(Z,q)F'(L)q + p(Z,q)F'(L) - w + \lambda_L = 0 \Rightarrow p_2(Z,q)F'(L)q + p(Z,q)F'(L) = w$

Differentiating this expression with respect to time and rearranging yields:

$$\dot{L} = -\frac{F'(L)[p_{12}(Z,q)q + p_1(Z,q)]}{F''(L)[p_2(Z,q)q + p(Z,q)] + F'(L)^2[p_{22}(Z,q)q + 2p_2(Z,q)]}\dot{Z}$$

$F'(L) > 0$, $F''(L)[p_2(Z,q)q + p(Z,q)] + F'(L)^2[p_{22}(Z,q)q + 2p_2(Z,q)] < 0$ and $\dot{Z} > 0$

$\Rightarrow Sign[\dot{L}] = Sign[p_{12}(Z,q)q + p_1(Z,q)]$

$p_1(Z,q) > 0$ and $Sign[\dot{L}] = Sign[p_{12}(Z,q)q + p_1(Z,q)] \Rightarrow p_{12}(Z,q) \geq 0$ is sufficient for $\dot{L} > 0$ and $p_{12}(Z,q) < 0$ is necessary for $\dot{L} \leq 0$

(R14) The employment of exploratory researchers decreases over time $\dot{N}_E < 0$

Differentiating $\pi_S = \frac{w_E}{g'(N_E)}$ with respect to time yields $\dot{\pi}_S = -w_E \frac{g''(N_E)}{g'(N_E)^2}\dot{N}_E$

Equating $\dot{\pi}_S = -w_E \frac{g''(N_E)}{g'(N_E)^2}\dot{N}_E$ and $\dot{\pi}_S = -\pi_Z h_2(N_A, S)$ and rearranging:

$$\dot{N}_E = \frac{g'(N_E)^2 \pi_Z h_2(N_A, S)}{g''(N_E) w_E}$$

$g'(N_E)^2 > 0$, $g''(N_E) \leq 0$, $\pi_Z > 0$, $h_2(N_A, S) > 0$ and $w_E > 0 \Rightarrow \dot{N}_E < 0$ [11]

(R15) The sign of $\dot{N}_A$ is undetermined; there is a positive (but not constant) upper bound to $\dot{N}_A$ positively related to the growth in knowledge resources $\dot{S}$

Differentiating $\pi_Z = \frac{w_A}{h_1(N_A, S)}$ with respect to time yields

$$\dot{\pi}_Z = -w_A \frac{[h_{11}(N_A, S)\dot{N}_A + h_{12}(N_A, S)\dot{S}]}{h_1(N_A, S)^2}$$

[11] Note that in the limit where $g''(N_E) = 0$, $\dot{N}_E$ is undefined, so it would be mathematically more correct to condition this results to a strictly concave equation of motion for knowledge resources.

Equating $\dot{\pi}_Z = -p_1(Z,q)q$ with $\dot{\pi}_Z = -w_A \dfrac{[h_{11}(N_A,S)\dot{N}_A + h_{12}(N_A,S)\dot{S}]}{h_1(N_A,S)^2}$ and rearranging:

$$\dot{N}_A = \frac{-1}{h_{11}(N_A,S)}\left[h_{12}(N_A,S)\dot{S} - \frac{p_1(Z,q)qh_1(N_A,S)^2}{w_A}\right]$$

$p_1(Z,q)>0$, $q>0$, $h_1(N_A,S)^2>0$ and $w_A>0 \Rightarrow \dot{N}_A < -\dfrac{h_{12}(N_A,S)}{h_{11}(N_A,S)}\dot{S}$,

where $h_{11}(N_A,S)<0$, $h_{12}(N_A,S)>0$ and $\dot{S}>0$

(R16) The state of fundamental knowledge increases at a decreasing rate $\ddot{S}<0$

Differentiating $\dot{S} = g(N_E)$ with respect to time yields $\ddot{S} = g'(N_E)\dot{N}_E$

$g'(N_E)>0$, $\dot{N}_E<0$ and $\ddot{S} = g'(N_E)\dot{N}_E \Rightarrow \ddot{S}<0$

(R17) $\dot{N}_A \geq 0$ is sufficient for quality improvements to accelerate over time $\ddot{Z}>0$; $\dot{N}_A<0$ is necessary but not sufficient for $\ddot{Z}\leq 0$; $\ddot{Z}\leq 0$ is sufficient for $\dot{N}_A<0$

The differential of $\dot{Z} = h(N_A,S)$ with respect to time is $\ddot{Z} = h_1(N_A,S)\dot{N}_A + h_2(N_A,S)\dot{S}$

$h_1(N_A,S)>0$, $h_2(N_A,S)>0$, $\dot{S}>0$ and $\ddot{Z} = h_1(N_A,S)\dot{N}_A + h_2(N_A,S)\dot{S} \Rightarrow \dot{N}_A \geq 0$ is sufficient for $\ddot{Z}>0$, $\dot{N}_A<0$ is necessary for $\ddot{Z}\leq 0$ and $\ddot{Z}\leq 0$ is sufficient for $\dot{N}_A<0$

(R18) A constant or accelerating decline in the shadow price of additional knowledge $\ddot{\pi}_S \leq 0$ is sufficient for applications-oriented research to be growing $\dot{N}_A>0$; $\dot{N}_A \leq 0$ is sufficient for $\ddot{\pi}_S>0$; $\dot{N}_A>0$ is necessary but not sufficient for $\ddot{\pi}_S \leq 0$

Differentiating $\dot{\pi}_S = -\pi_Z h_2(N_A,S)$ with respect to time gives:

$$\ddot{\pi}_S = -\dot{\pi}_Z h_2(N_A,S) - \pi_Z h_{12}(N_A,S)\dot{N}_A - \pi_Z h_{22}(N_A,S)\dot{S}$$

$\dot{\pi}_Z < 0$, $h_2(N_A,S) > 0$, $\pi_Z > 0$, $h_{12}(N_A,S) > 0$, $h_{22}(N_A,S) < 0$ and $\dot{S} > 0 \Rightarrow \ddot{\pi}_S \leq 0$ is sufficient for $\dot{N}_A > 0$, $\dot{N}_A \leq 0$ is sufficient for $\ddot{\pi}_S > 0$ and $\dot{N}_A > 0$ is necessary for $\ddot{\pi}_S \leq 0$

(R19) If $h(N_A,S)$ is homogeneous of degree 1, the decline in the shadow price of additional knowledge is decelerating $\ddot{\pi}_S > 0$

$h(N_A,S)$ homogeneous of degree 1 $\Rightarrow h(N_A,S) = h_1(N_A,S)N_A + h_2(N_A,S)S$

Differentiating the preceding expression with respect to time and rearranging gives:

$$h_{12}(N_A,S)\dot{N}_A + h_{22}(N_A,S)\dot{S} = -\frac{N_A}{S}\left[h_{11}(N_A,S)\dot{N}_A + h_{12}(N_A,S)\dot{S}\right]$$

$N_A > 0$, $S > 0$ and $h_{11}(N_A,S)\dot{N}_A + h_{12}(N_A,S)\dot{S} > 0$ from (R15)

$\Rightarrow h_{12}(N_A,S)\dot{N}_A + h_{22}(N_A,S)\dot{S} < 0$

$\dot{\pi}_Z < 0$, $h_2(N_A,S) > 0$, $\pi_Z > 0$ and $h_{12}(N_A,S)\dot{N}_A + h_{22}(N_A,S)\dot{S} < 0$

$\Rightarrow \ddot{\pi}_S = -\dot{\pi}_Z h_2(N_A,S) - \pi_Z\left[h_{12}(N_A,S)\dot{N}_A + h_{22}(N_A,S)\dot{S}\right] > 0$

(R20) The decline in the shadow price of quality improvements is decelerating $\ddot{\pi}_Z > 0$

Differentiating $\dot{\pi}_Z = -p_1(Z,q)q$ with respect to time leads to the following:

$$\ddot{\pi}_Z = -p_{11}(Z,q)q\dot{Z} - p_{12}(Z,q)qF'(L)\dot{L} - p_1(Z,q)F'(L)\dot{L}$$

Substituting for $\dot{L} = -\dfrac{F'(L)[p_{12}(Z,q)q + p_1(Z,q)]}{F''(L)[p_2(Z,q)q + p(Z,q)] + F'(L)^2[p_{22}(Z,q)q + 2p_2(Z,q)]}\dot{Z}$,

$$\ddot{\pi}_Z = \left[-p_{11}(Z,q)q + \frac{F'(L)^2[p_{12}(Z,q)q + p_1(Z,q)]^2}{F''(L)[p_2(Z,q)q + p(Z,q)] + F'(L)^2[p_{22}(Z,q)q + 2p_2(Z,q)]}\right]\dot{Z}$$

$\dot{Z} > 0$ and from the conditions of concavity of the revenue function,

$$-p_{11}(Z,q)q + \frac{F'(L)^2[p_{12}(Z,q)q + p_1(Z,q)]^2}{F''(L)[p_2(Z,q)q + p(Z,q)] + F'(L)[p_{22}(Z,q)q + 2p_2(Z,q)]} > 0$$

$\Rightarrow \ddot{\pi}_Z > 0$

Table 3.1 Summary of results for product quality oriented innovation

'Anticipated' results	'Unanticipated' results
(R1) $\lambda_L = 0$	(R8) $\dot{q} \leq 0$ is sufficient for $\dot{p} > 0$
$\lambda_{N_A} = 0$	$\dot{q} > 0$ is necessary for $\dot{p} \leq 0$
$\lambda_{N_E} = 0$	$\dot{p} \leq 0$ is sufficient for $\dot{q} > 0$
(R2) $q > 0$	(R10) $\dot{L} \leq 0$ is sufficient for $\dot{p} > 0$
(R3) $p > 0$	$\dot{L} > 0$ is necessary for $\dot{p} \leq 0$
(R4) $\pi_S > 0$	$\dot{p} \leq 0$ is sufficient for $\dot{L} > 0$
(R5) $\pi_Z > 0$	(R11) $\dot{\pi}_S < 0$
(R6) $\dot{S} > 0$	(R12) $\dot{\pi}_Z < 0$
(R7) $\dot{Z} > 0$	(R13) $Sign[\dot{L}] = Sign[p_{12}(Z,q)q + p_1(Z,q)]$
(R9) $Sign[\dot{q}] = Sign[\dot{L}]$	$p_{12}(Z,q) \geq 0$ is sufficient for $\dot{L} > 0$
	$p_{12}(Z,q) < 0$ is necessary for $\dot{L} \leq 0$
	(R14) $\dot{N}_E < 0$
	(R15) $\dot{N}_A$ any sign
	$\dot{N}_A < -\frac{h_{12}(N_A,S)}{h_{11}(N_A,S)}\dot{S}$
	(R16) $\ddot{S} < 0$
	(R17) $\dot{N}_A \geq 0$ is sufficient for $\ddot{Z} > 0$
	$\dot{N}_A < 0$ is necessary for $\ddot{Z} \leq 0$
	$\ddot{Z} \leq 0$ is sufficient for $\dot{N}_A < 0$
	(R18) $\ddot{\pi}_S \leq 0$ is sufficient for $\dot{N}_A > 0$
	$\dot{N}_A \leq 0$ is sufficient for $\ddot{\pi}_S > 0$
	$\dot{N}_A > 0$ is necessary for $\ddot{\pi}_S \leq 0$
	(R19) If $h(N_A,S)$ is homogeneous of degree 1 then $\ddot{\pi}_S > 0$
	(R20) $\ddot{\pi}_Z > 0$

Table 3.1 presents a summary of the results for the case of product innovation oriented research, divided into two categories. The first category is that of results following immediately from the assumption of an interior solution and which are therefore to be

'anticipated.' (R1) to (R7) and (R9) are placed into this category, since they depend most directly on L, N_A and N_E being positive. Results (R8) and (R10) to (R20) are placed in the category of 'unanticipated' findings, in the sense of not being immediately forthcoming from these assumptions, but derived from the various necessary conditions through nontrivial mathematical manipulations. This is a heuristic division and some may argue that one should place more (even many more) of the results in the first column. Be that as it may, it serves to focus attention on the properties listed in the second column, the implications of which are discussed here and in the concluding section of this chapter.

A description of these findings may begin by drawing attention to result (R14), that a firm engaged in production, exploratory- and applications-oriented research should reduce its expenditure on exploratory research over time, as a consequence of the firm's decreasing valuation of fundamental knowledge additions (R11) and concavity of the production function for this class of knowledge resources. Accordingly, we can be more precise about the dynamic evolution of fundamental knowledge resources available to applications-oriented researcher workers than simply to state that it is growing (R6): the 'optimised' state of knowledge would grow at a decreasing rate (R16).

Even though a monopolist also would impute a lower value to the practical knowledge embodied in quality improvements as it gets closer to the end of the planning horizon (R12), the effort to create an improved product may intensify, stay the same or weaken over time (R15). The order in which this occurs cannot be proven formally without additional assumptions. The difficulty in determining the optimal rate of change of applications-oriented research is that additional knowledge raises the marginal productivity of this mode of research, but the imputed value to additional quality declines over time, depressing the marginal productivity of additions to quality in profit terms; only if these two 'forces' exactly cancel out can there be a steady state in N_A. The growth of applications-

oriented research is nonetheless bounded from above by the growth in fundamental knowledge. Therefore, whilst product quality is obviously growing in an interior solution (R7), it will do so more rapidly over time if there is a constant or increasing level of resources devoted to applications research (R17). In the final section we will argue that the set of results dealing with the optimal research plan can generate a 'product life cycle.'

Turning to the results relating to the production plan, we gain several (potentially testable) implications. To begin with, there are rather trivial predictions about a positive correlation between labour inputs and output (R9) and a negative one between price and output, which follow directly from the assumptions about the inverse demand and the production function. More interestingly, (R10) shows that a firm whose quality is increasing should be increasing its employment in manufacturing if the price is constant or decreasing; if there is a steady state in or falling employment, the price set should be rising over time. The latter could be interpreted as a strategy of increasing quality while making the good more exclusive by cutting production and simultaneously raising its price. Finally, if the cross elasticity of demand is such that $p_{12}(Z,q) \geq 0$, the optimal dynamic path of labour inputs and output is exactly pinned down, both should be increasing.

As a final point, one can consider the implications of the solution's properties for a firm employing a fixed labour force, i.e., one in which the total number of workers allocated to manufacturing and research was unchanging for a period of time. If $\dot{L} = 0$ and $\dot{N}_E + \dot{N}_A = 0$, then $\dot{p} > 0$, $\dot{N}_A > 0$, $\ddot{Z} > 0$ and $\ddot{\pi}_S > 0$. That is to say, for a monopolist's production and research facilities to be profit maximising when their scale is fixed, the price set will need to rise over time and the effort of its researchers should be progressively redirected from expanding the fundamental knowledge base towards improving its product design. The increasing expenditure on applications-oriented research would lead to an accelerating growth in product quality.

3.4 Own-process innovation

The problem of designing an optimal production and research programme when the resources are allocated to improving the technology of production has many similarities with the case of innovations in product quality. Consequently, this section is sparser in showing how the results are derived, but the methodology for doing so is the same as in Section 3.3. In terms of the model set-up, the differences are that: product quality is fixed – to emphasise this fact and simplify notation quality is set to 1 in the inverse demand function $p(1,q)$; output is a function of the varying state of the production technology $q = TF(L)$; and applications-oriented researchers now work on improving the efficiency of the manufacturing processes $\dot{T} = h(N_A, S)$. The Hamiltonian function for this problem is:

$$H = p(1, TF(L))TF(L) - wL - w_A N_A - w_E N_E + \pi_T h(N_A, S) + \pi_S g(N_E) \qquad (3.13)$$

where the costate variable corresponding to the differential equation $\dot{T}$, denoted as π_T, is a measure of the imputed value in profit terms of increasing the production technology state. The Lagrangian function is:

$$\begin{aligned} \mathsf{L} &= H + \lambda_L L + \lambda_{N_A} N_A + \lambda_{N_E} N_E \\ &= p(1, TF(L))TF(L) - wL - w_A N_A - w_E N_E + \pi_T h(N_A, S) + \pi_S g(N_E) \\ &\quad + \lambda_L L + \lambda_{N_A} N_A + \lambda_{N_E} N_E \end{aligned} \qquad (3.14)$$

The necessary conditions for an optimal solution can be derived from the Lagrangian:

$$\frac{\partial \mathsf{L}^*}{\partial L} = p_2(1,q)TF'(L)q + p(1,q)TF'(L) - w + \lambda_L = 0 \qquad (3.15)$$

$$\frac{\partial \mathsf{L}^*}{\partial N_A} = -w_A + \pi_T h_1(N_A, S) + \lambda_{N_A} = 0 \qquad (3.16)$$

$$\frac{\partial \mathsf{L}^*}{\partial N_E} = -w_E + \pi_S g'(N_E) + \lambda_{N_E} = 0 \tag{3.17}$$

$$\lambda_L \geq 0 \qquad L \geq 0 \qquad \lambda_L L = 0$$

$$\lambda_{N_A} \geq 0 \qquad N_A \geq 0 \qquad \lambda_{N_A} N_A = 0 \tag{3.18}$$

$$\lambda_{N_E} \geq 0 \qquad N_E \geq 0 \qquad \lambda_{N_E} N_E = 0$$

$$\dot{\pi}_T = -\frac{\partial \mathsf{L}^*}{\partial T} = -p_2(1,q)F(L)q - p(1,q)F(L) \tag{3.19}$$

$$\dot{\pi}_S = -\frac{\partial \mathsf{L}^*}{\partial S} = -\pi_T h_2(N_A, S) \tag{3.20}$$

$$\dot{T} = \frac{\partial \mathsf{L}^*}{\partial \pi_T} = h(N_A, S) \tag{3.21}$$

$$\dot{S} = \frac{\partial \mathsf{L}^*}{\partial \pi_S} = g(N_E) \tag{3.22}$$

$$T(t_0) = T_0 > 0 \qquad \pi_T(t_1) = 0 \tag{3.23}$$

$$S(t_0) = S_0 > 0 \qquad \pi_S(t_1) = 0 \tag{3.24}$$

Plainly, the parallels in the planning decisions for product quality and own-process innovations lead to a very similar set of necessary conditions. In effect, only two equations are appreciably different. One is the first order condition with respect to production labour in (3.15), which now establishes a correspondence between the state of the production technology, the size of the labour force and its wage that must hold at each point in time. The optimisation of the labour input requirements must factor in the influence of the production technology on the marginal productivity of labour. Notice that in an interior solution

$p_2(1,q)q + p(1,q) \leq 0$ leads to a contradiction. We assume, therefore, that $p_2(1,q)q + p(1,q) > 0$ in what remains of this section.

The other condition that has changed is the one setting the optimal change in the costate variable associated to production technology improvements (equation 3.19). In this setting, it weighs the marginal revenue product with respect to the state of the firm's production efficiency:

$$\dot{\pi}_T = -F(L)[p_2(1,q)q + p(1,q)] = -\frac{\partial[p(1,q)q]}{\partial T}$$

The sufficient conditions for an optimum in the case of process innovations are (as in the case of product innovation) that the objective functional, the two equations of motion and the inequality constraints are jointly concave in the state and control variables. As $h(N_A, S)$, $g(N_E)$ and the nonnegativity constraints satisfy these requirements under the conditions outlined in the 'model' with product innovation, the sufficient conditions are fulfilled whenever the Hessian of the revenue function is negative semidefinite. Let us assume once again that the revenue function is strictly concave, in which case its Hessian needs to be negative definite. This calls for the second derivative of the revenue with respect to production technology improvements to be less than zero,

$$F(L)^2[p_{22}(1,q)TF(L) + 2p_2(1,q)] < 0,$$

which implies that $p_{22}(1,q)q + 2p_2(1,q) < 0$ in an interior solution. In addition, the determinant of the Hessian matrix needs to be greater than zero:

$$F(L)^2[p_{22}(1,q)q + 2p_2(1,q)]\left\langle T^2F'(L)^2[p_{22}(1,q)q + 2p_2(1,q)] + TF''(L)[p_2(1,q)q + p(1,q)]\right\rangle$$

$$-F'(L)^2[p_{22}(1,q)q^2 + 3p_2(1,q)q + p(1,q)]^2 > 0$$

Or, rearranging:

$$T^2F'(L)^2\left[p_{22}(1,q)q+2p_2(1,q)\right]+TF''\left[p_2(1,q)q+p(1,q)\right]<$$

$$\frac{F'(L)^2\left[p_{22}(1,q)q^2+3p_2(1,q)q+p(1,q)\right]^2}{F(L)^2\left[p_{22}(1,q)q+2p_2(1,q)\right]}\leq 0$$

This intricate expression can be simplified further by noting that:

$$p_{22}(1,q)q^2+3p_2(1,q)q+p(1,q)=\left[p_{22}(1,q)q+2p_2(1,q)\right]q+\left[p_2(1,q)q+p(1,q)\right]$$

By rearranging one can obtain, after some algebraic manipulations, the following condition:

$$F''(L)\leq\frac{F'(L)^2}{F(L)}\left[2+\frac{p_2(1,q)q+p(1,q)}{\left[p_{22}(1,q)q+2p_2(1,q)\right]q}\right]$$

Therefore concavity of the intermediate function imposes: (1) a positive upper bound on the second derivative of the inverse demand function with respect to output, $p_{22}(1,q)\leq 0$ is sufficient for the assumption to be satisfied, but not necessary; and (2) an upper bound to $F''(L)$. The later may or may not be binding, i.e., concavity of the production function could be sufficient for condition (2) to hold, it will depend on the precise value of the expression in square brackets. The sense of these conditions is that the inverse demand function and the production function must be sufficiently concave in their arguments for the revenue function to be jointly concave in T and L. The nature of the restrictions will become clearer in Section 3.5, as specific functional forms allow the sufficient conditions to be posed in terms of the elasticities of demand and production.

In an interior solution, the majority of the properties derived earlier also hold for process innovation; this is true, in particular, of all the properties pertaining to the optimised dynamic profile of research. It can be shown that: $\lambda_L=0$, $\lambda_{N_A}=0$, $\lambda_{N_E}=0$, $q>0$,

$p = p(1,q) > 0$, $\pi_S > 0$, $\pi_T > 0$, $\dot{S} > 0$, $\dot{T} > 0$, $Sign[\dot{p}] = -Sign[\dot{q}]$, $\dot{\pi}_S < 0$, $\dot{\pi}_T < 0$, $\dot{N}_E < 0$, $\ddot{S} < 0$, if $h(N_A, S)$ is homogeneous of degree 1 then $\ddot{\pi}_S > 0$, and $\ddot{\pi}_T > 0$. $\dot{N}_A$ can be greater than, less than, or equal to zero, and the relationships between this rate of change and "acceleration" or "deceleration" in the state of the production technology $\ddot{T} > 0$ and the shadow price of knowledge increases $\ddot{\pi}_S$ are the same as with product innovation. The following results determine the properties that must be satisfied by the optimal trajectories for labour demand, quantities produced and price (the numbers correspond to the order of the results for the analysis of product quality), the only ones that are qualitatively different:

(R8*) The rate of change of price has the opposite sign as the rate of change of output $Sign[\dot{p}] = -Sign[\dot{q}]$

Differentiating $p = p(1,q)$ with respect to time gives $\dot{p} = p_2(1,q)\dot{q}$

$p_2(1,q) < 0$ and $\dot{p} = p_2(1,q)\dot{q} \Rightarrow Sign[\dot{p}] = -Sign[\dot{q}]$

(R9*) A constant or growing employment in manufacturing $\dot{L} \geq 0$ is sufficient for output to grow over time $\dot{q} > 0$; $\dot{L} < 0$ is necessary but not sufficient for $\dot{q} \leq 0$; $\dot{q} \leq 0$ is sufficient for $\dot{L} < 0$

Differentiating $q = TF(L)$ with respect to time gives $\dot{q} = \dot{T}F(L) + TF'(L)\dot{L}$

$T > 0$, $\dot{T} > 0$, $F(L) > 0$, $F'(L) > 0$ and $\dot{q} = \dot{T}F(L) + TF'(L)\dot{L} \Rightarrow \dot{L} \geq 0$ is sufficient for $\dot{q} > 0$, $\dot{L} < 0$ is necessary for $\dot{q} \leq 0$ and $\dot{q} \leq 0$ is sufficient for $\dot{L} < 0$

(R10*) A constant or growing employment in manufacturing $\dot{L} \geq 0$ is sufficient for the price to decrease over time $\dot{p} < 0$; $\dot{L} < 0$ is necessary but not sufficient for $\dot{p} \geq 0$; $\dot{p} \geq 0$ is sufficient for $\dot{L} < 0$

Differentiating $p = p(1,TF(L))$ with respect to time gives $\dot{p} = p_2(1,q)\left[\dot{T}F(L) + TF'(L)\dot{L}\right]$

$p_2(1,q) < 0$, $\dot{T} > 0$, $F(L) > 0$, $T > 0$, $F'(L) > 0$ and $\dot{p} = p_2(1,q)\left[\dot{T}F(L) + TF'(L)\dot{L}\right] \Rightarrow \dot{L} \geq 0$

is sufficient for $\dot{p} < 0$, $\dot{L} < 0$ is necessary for $\dot{p} \geq 0$ and $\dot{p} \geq 0$ is sufficient for $\dot{L} < 0$

(R13*) The sign of $\dot{L}$ is the same as that of $p_{22}(1,q)q^2 + 3p_2(1,q)q + p(1,q)$

$\lambda_L = 0$ and $p_2(1,q)TF'(L)q + p(1,q)TF'(L) - w + \lambda_L = 0$

$\Rightarrow p_2(1,q)TF'(L)q + p(1,q)TF'(L) = w$

Differentiating this expression with respect to time and rearranging leads to:

$$\dot{L} = -\frac{F'(L)\left[p_{22}(1,q)q^2 + 3p_2(1,q)q + p(1,q)\right]}{T^2F'(L)^2\left[p_{22}(1,q)q + 2p_2(1,q)\right] + TF''(L)\left[p_2(1,q)q + p(1,q)\right]}\dot{T}$$

$F'(L) > 0$, $T^2F'(L)^2\left[p_{22}(1,q)q + 2p_2(1,q)\right] + TF''(L)\left[p_2(1,q)q + p(1,q)\right] < 0$ and $\dot{T} > 0$

$\Rightarrow Sign\left[\dot{L}\right] = Sign\left[p_{22}(1,q)q^2 + 3p_2(1,q)q + p(1,q)\right]$

As in the previous section, Table 3.2 presents a categorised summary of the results for the case of process innovation oriented research. It indicates that spending on exploratory research again should be cut down over time (R14*). Similarly, spending on applications-oriented research can be increasing, constant or decreasing (R15*). These optimal controls imply that the state of knowledge is growing at a decreasing rate (R16*), but process efficiency can rise more or less rapidly over time depending on the profile of resources devoted to applications-oriented research (R17*). In short, the R&D variables associated with process innovation have the same properties, even if they are not connected in exactly the same way as before. I.e., the evolution of the control, state and costate variables no longer reflects the responsiveness of demand to changes in product quality; instead, the focus is on the responsiveness to price changes, and the properties of the production function.

Table 3.2 Summary of results for own-process oriented research

'Anticipated' results	'Unanticipated' results
(R1*) $\lambda_L = 0$ $\lambda_{N_A} = 0$ $\lambda_{N_E} = 0$	(R9*) $\dot{L} \geq 0$ is sufficient for $\dot{q} > 0$ $\dot{L} < 0$ is necessary for $\dot{q} \leq 0$ $\dot{q} \leq 0$ is sufficient for $\dot{L} < 0$
(R2*) $q > 0$	(R10*) $\dot{L} \geq 0$ is sufficient for $\dot{p} < 0$ $\dot{L} < 0$ is necessary for $\dot{p} \geq 0$ $\dot{p} \geq 0$ is sufficient for $\dot{L} < 0$
(R3*) $p > 0$	(R11*) $\dot{\pi}_S < 0$
(R4*) $\pi_S > 0$	(R12*) $\dot{\pi}_T < 0$
(R5*) $\pi_T > 0$	(R13*) $Sign[\dot{L}] = Sign[p_{22}(1,q)q^2 + 3p_2(1,q)q + p(1,q)]$
(R6*) $\dot{S} > 0$	(R14*) $\dot{N}_E < 0$
(R7*) $\dot{T} > 0$	(R15*) $\dot{N}_A$ any sign $\dot{N}_A < -\frac{h_{12}(N_A,S)}{h_{11}(N_A,S)}\dot{S}$
(R8*) $Sign[\dot{p}] = -Sign[\dot{q}]$	(R16*) $\ddot{S} < 0$
	(R17*) $\dot{N}_A \geq 0$ is sufficient for $\ddot{T} > 0$ $\dot{N}_A < 0$ is necessary for $\ddot{T} \leq 0$ $\ddot{T} \leq 0$ is sufficient for $\dot{N}_A < 0$
	(R18*) $\ddot{\pi}_S \leq 0$ is sufficient for $\dot{N}_A > 0$ $\dot{N}_A \leq 0$ is sufficient for $\ddot{\pi}_S > 0$ $\dot{N}_A > 0$ is necessary for $\ddot{\pi}_S \leq 0$
	(R19*) If $h(N_A,S)$ is homogeneous of degree 1 then $\ddot{\pi}_S > 0$
	(R20*) $\ddot{\pi}_T > 0$

The main difference between a firm specialising its applications-oriented research in product innovation and one orienting its research efforts towards process innovation concerns the optimised production plan. In the present case, optimal price and output are positively related (R8*). In contrast, the relationship between employment in manufacturing and output (R9*) is no longer as direct because the marginal productivity of labour is increasing as a

consequence of process innovations: output tends to grow because the production technology becomes more efficient, but this will generally entail concurrent layoffs in the workforce to cut the wage bill of workers in manufacturing and moderate the fall in the willingness to pay that takes place if the firm simply 'floods' the market. An optimal pattern of employment, production and pricing in a process innovating firm emerges from (R9*) and (R10*): if employment in manufacturing is constant or additional workers are being hired by the firm, its output would be increasing and the price it sets decreasing over time; if the price is to remain rigid or to increase, the firm must layoff workers over time.

In the hypothetical case of a process-innovating firm with a constant number of workers allocated to manufacturing and research (that is, when there is a steady state in L and $N_A + N_E$ for a certain time), a profit-maximising plan would be consistent with: increasing output; a declining price; and a reallocation of researchers from exploratory to applications oriented research over time. Result (13*) can be drawn on to clarify the conditions under which the firm would want to have a fixed scale of production. For a steady state in the demand for labour ($\dot{L} = 0$) to be optimal, the marginal revenue product of labour must remain unchanged in response to a more efficient production technology. To understand why this is the case, note that the change in the marginal revenue product of labour as the production technology becomes marginally more efficient equals:

$$\frac{\partial}{\partial T}\left[\frac{\partial[p(1,q)q]}{\partial L}\right] = \frac{\partial^2[p(1,q)q]}{\partial T \partial L} = F'(L)\left[p_{22}(1,q)q^2 + 3p_2(1,q) + p(1,q)\right]$$

3.5 Example with specific functional forms

Solving the model for specific (albeit purposefully flexible) functional forms is useful because it: (i) spells out the meaning of some of the assumptions made, notably the

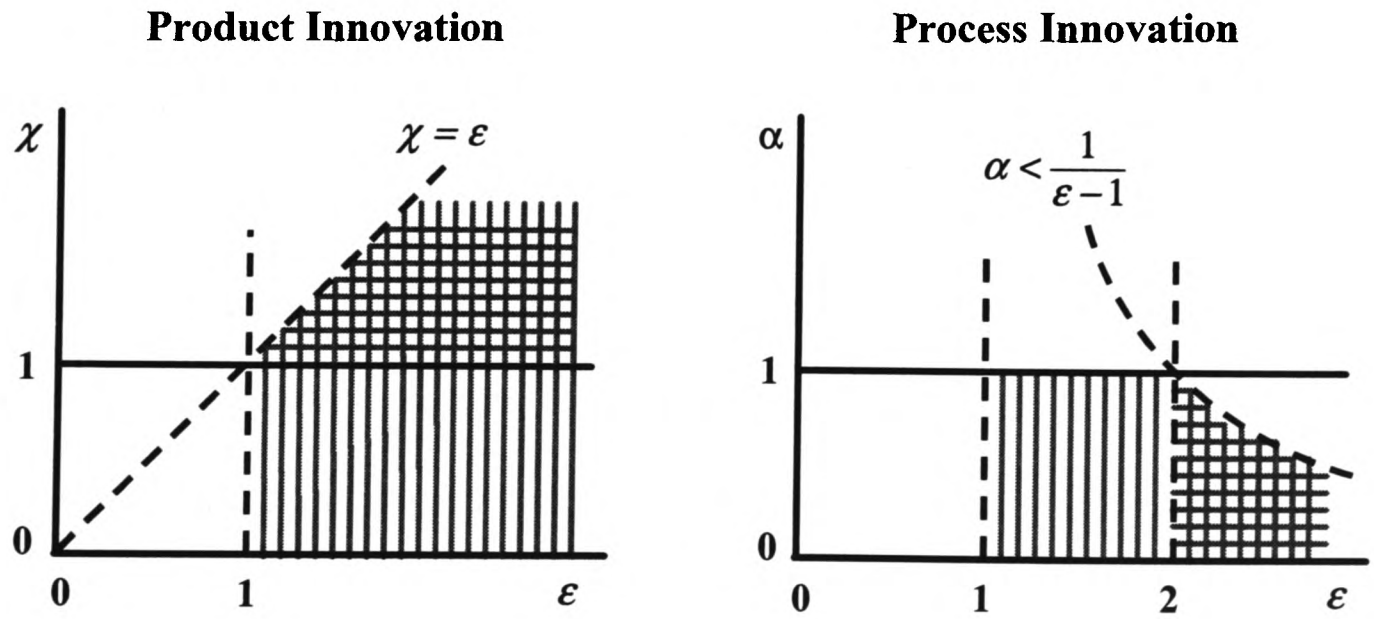
complicated expressions for concavity, in terms of recognisable parameters such as the elasticity of demand with respect to price or the share of research workers in inventing improved product designs; (ii) leads to simple relationships between the optimised growth rates of the state and control variables, which help to give a 'feel' for the magnitudes of the growth rates of the control variables; and (iii) determines the directions of some of the trajectories that had ambiguous paths in the general formulation. This section considers each of these three objectives in turn. As an aside, note that this is exactly the specification employed in the following chapter to analyse the dynamic problem using numerical techniques – concerning this inquiry, more anon, in Section 3.6.

We consider the following functional forms. (1) The inverse demand function is specified as $p(Z,q)=Z^{\chi/\varepsilon}q^{-1/\varepsilon}$, where $\varepsilon>0$ is the elasticity of demand with respect to price and $\chi>0$ is the elasticity of demand with respect to product quality, as can readily be observed from the corresponding Marshallian demand function $q(Z,p)=Z^{\chi}p^{-\varepsilon}$. (2) Let the firm's production function be $q=TL^{\alpha}$, where $0<\alpha\leq 1$. (3) The evolution of product quality is $\dot{Z}=N_A{}^{\nu}S^{1-\nu}$, as is that of production technology $\dot{T}=N_A{}^{\nu}S^{1-\nu}$; ν is constant with $0<\nu<1$, where ν is the share of research labour in applications-oriented research, be it for improved products or processes, and $1-\nu$ is the share of scientific knowledge in advancing product or production system designs. (4) The evolution of the knowledge state is $\dot{S}=N_E{}^{\phi}$, with ϕ constant with $0<\phi<1$. With the parameter values specified, these functional forms meet all of the assumptions (about continuity, concavity, etcetera) made in respect to the general functional relations; excluding $\phi=1$ avoids the limit case in which the rate of change for exploratory research workers was undefined.

It remains to check what parameter subspace ensures strict concavity of the revenue function. In the case of product innovation, the first condition for strict concavity simplifies

to $\varepsilon > \chi$: the price elasticity of demand must be greater than the elasticity of demand with respect to product quality. The second condition, concerning the determinant of the Hessian matrix, requires that $\varepsilon > 1$ and $\alpha < [\varepsilon - \chi]/[\varepsilon - 1]$. Notice that the upper bound on α is redundant if $0 < \chi < 1$, given that this parameter needs to be constrained to $0 < \alpha \leq 1$ anyhow, but binding whenever $\chi \geq 1$. When we turn to improvements in the production system the conditions are even simpler, requiring the price elasticity of demand to be greater than one, $\varepsilon > 1$, and that the parameter in the production function satisfy $\alpha < 1/[\varepsilon - 1]$. If the price elasticity of demand is in the open interval $1 < \varepsilon < 2$, the latter condition is redundant. This has been depicted in Figure 3.1; the blank area in the diagrams is not admissible, in the area with vertical lines the second condition is redundant, and in the criss-crossed area it is binding. Finally, notice that the additional assumption that was needed for the necessary conditions in the interior solution of the model of process innovation to avoid a contradiction, $p_2(1,q)q + p(1,q) > 0$, also calls for $\varepsilon > 1$.

Figure 3.1 Sufficient conditions for a global maximum



Whatever the orientation of research, it appears that an elastic demand ($\varepsilon > 1$) is a central part of the sufficiency conditions for a global maximum. The meaning of this restriction in

conventional economic terms is that, whilst a monopolist would like to operate in the inelastic segment of the demand curve (where $\varepsilon < 1$), when ε is in the interval $(0,1)$ revenue is an increasing function of the price. This can be observed from the revenue function constructed using the Marshallian demand, which equals $pq(Z,p) = Z^{\chi} p^{1-\varepsilon}$. This convexity precludes a solution to the maximisation problem. When $\varepsilon = 1$, revenue is independent of the price, and the optimal employment in manufacturing is undefined. Imposing a minimum degree of price elasticity may seem an overly restrictive assumption and, in fact, it is. One needs to be mindful that this is a sufficient condition, not a necessary one. Let it be noted in passing that the optimal control problem could support such convexities in revenue, or indeed in the equations of motion, but then there must be "enough" concavity in another part of the dynamic system for a solution to exist. For instance, if the firm faced a less than perfectly elastic labour supply, the requirement for the Lagrangian to be concave could be met with a less stringent restriction on the elasticities pertaining to the product market.[12]

The 'trade-off' between the concavity in one part of the problem and another underlies the other sufficiency conditions as well. In particular, the condition that $\alpha < [\varepsilon - \chi]/[\varepsilon - 1]$ for product innovation indicates that when demand is not very responsive to changes in quality, all that is needed for the control problem to be well defined is an inelastic price elasticity of demand and a concave production function. If the elasticity of demand with respect to quality is higher, demand must be more sensitive to the price and there must be significant decreasing returns to labour to prevent convexity in the revenue function. Some numbers make the same point: if $\varepsilon = 1.25$ and $\chi = 0.75$, then $\alpha < 2$, which

[12] Arrow's seminal article of 1962, which analysed the allocation of R&D resources, is relevant in this respect. In the static analysis of his paper, a monopolist *can* determine its equilibrium spending in R&D in the inelastic segment of demand because the reduction of marginal cost brought about by investment in R&D is strictly concave. This is not possible in our model because the firm has constant marginal (labour) costs of production; and, because our problem is dynamic in nature, there are extra constraints that need to be fulfilled.

is always true; if $\varepsilon = 1.25$ and $\chi = 1.1$, then $\alpha < .6$. The same argument applies to process innovation, where $\alpha < 1/[\varepsilon - 1]$ is redundant so long as $1 < \varepsilon < 2$, but when demand is very elastic to price there must be decreasing returns to scale for revenue to be concave – otherwise, a firm could get unlimited profits by cutting its price and increasing output.

Table 3.3 Optimised growth rates in a specified model

Product-oriented innovation	Process-oriented innovation
$\gamma_L = \frac{\chi}{\varepsilon[1-\alpha]+\alpha}\gamma_Z > 0$	$\gamma_L = \frac{\varepsilon - 1}{\varepsilon[1-\alpha]+\alpha}\gamma_T > 0$
$\gamma_{N_A} = \gamma_S + \frac{1}{1-\upsilon}\gamma_{\pi_Z}$	$\gamma_{N_A} = \gamma_S + \frac{1}{1-\upsilon}\gamma_{\pi_T}$
$\gamma_{N_E} = \frac{1}{1-\phi}\gamma_{\pi_S} < 0$	$\gamma_{N_E} = \frac{1}{1-\phi}\gamma_{\pi_S} < 0$
$\gamma_p = \frac{\chi}{\varepsilon}\gamma_Z - \frac{\alpha}{\varepsilon}\gamma_L$ $= [1-\alpha]\frac{\chi}{\varepsilon[1-\alpha]+\alpha}\gamma_Z > 0$	$\gamma_p = -\frac{1}{\varepsilon}\gamma_T - \frac{\alpha}{\varepsilon}\gamma_L$ $= -\frac{1}{\varepsilon[1-\alpha]+\alpha}\gamma_T < 0$
$\gamma_{pq} = \frac{\chi}{\varepsilon}\gamma_Z + \alpha\frac{\varepsilon - 1}{\varepsilon}\gamma_L$ $= \frac{\chi}{\varepsilon[1-\alpha]+\alpha}\gamma_Z > 0$	$\gamma_{pq} = \frac{\varepsilon - 1}{\varepsilon}\gamma_T + \alpha\frac{\varepsilon - 1}{\varepsilon}\gamma_L$ $= \frac{\varepsilon - 1}{\varepsilon[1-\alpha]+\alpha}\gamma_T > 0$
$\gamma_q = \alpha\gamma_L$ $= \alpha\frac{\chi}{\varepsilon[1-\alpha]+\alpha}\gamma_Z > 0$	$\gamma_q = \gamma_T + \alpha\gamma_L$ $= \frac{\varepsilon}{\varepsilon[1-\alpha]+\alpha}\gamma_T > 0$

Table 3.3 summarises the key mathematical relationships connecting the optimal growth rates of the control, state and costate variables for both improved product and production system designs; the growth rate of variable i is indicated by γ_i. The table shows, first of all, that the growth rates of each variable are constant proportions of each other. This permits a sharper

investigation into the nature of the optimal paths. Consider first the optimised growth rates that are common to both product- and process-oriented innovation. The rate of decline in exploratory research γ_{N_E} is greater than the rate of change in the imputed value of knowledge additions γ_{π_S}.[13] The rate of change of research for new applications γ_{N_A} is directly proportional to the rate of change in the state of knowledge, but decreases the faster is the (negative) rate of change of the imputed value of product quality. As a result, the requirement for a steady state of applications-driven research $\gamma_{N_A}=0$ is $[1-\upsilon]\gamma_S=-\gamma_{\pi_Z}$: the rate of growth of knowledge resources times the share of knowledge resources in eliciting product quality must equal the rate of change in the value imputed to product quality improvements. With process innovation, the relevant shadow price is γ_{π_T}.

Amongst the results specific to product innovation, one can note that the growth rate in output is a fraction of the growth rate in the labour force $\gamma_q=\alpha\gamma_L$; evidently, output does not increase as fast as employment in manufacturing because of diminishing returns to scale in production. The growth rate of price implied by the optimal controls is a weighted sum of the growth rates of product quality and of labour force employed in production; the latter exerts a downward influence because an increase in output lessens the willingness to pay by consumers. Substituting in for γ_L, it can be observed that the first effect dominates for these functional forms (because $p_{12}(Z,q)>0$). Therefore, the price set and the quantity produced by the firm would be growing over time, in step with the growth of product quality. Using the growth rates of output and price one can find the growth rate of revenue, which also is a fraction of the growth in product quality. The growth rate of revenue is closer to that of

[13] Since the functional form for the equations of motion describing the evolution of product quality or technical efficiency, $\dot{Z}=\dot{T}=h(N_A,S)=N_A{}^{\nu}S^{1-\nu}$, is homogeneous of degree 1, the fall in the valuation of (additional) knowledge resources will slow down as the terminal horizon is approached.

product quality the more elastic demand is to quality, the less elastic demand is to price, or the less pronounced are decreasing returns to scale in the production process.

In short, with these specific functional forms the growth rates of the demand for manufacturing labour, output produced, price and revenue are all positive in the case of product oriented innovation and hold precise relations that depend on the price and quality elasticity of demand and the other parameters. For process innovation, the firm will manage its 'research portfolio' and output so that quantities produced and revenue are increasing over time, and these decisions imply a decreasing price for the commodity sold. In this case the growth rate of revenue is closer to γ_T when demand is more responsive to price, or when output increases do not incur sharply diminishing returns to production. When the firm is improving its production technology, the optimal plan calls for a reduction in the price to increase the volume of goods marketed, thus making the most its lower production costs.

Introducing some hypothetical numbers can give an indication about the orders of magnitude by which the variables would be changing as a result of innovation. Suppose that:

- the elasticity of demand with respect to quality is $\chi = 0.75$,
- the price elasticity of demand is $\varepsilon = 1.25$,
- α in the production function is 0.8,
- quality is increasing by $\gamma_Z = 10\%$ per year, as is
- production efficiency growth $\gamma_T = 10\%$ per year.

In the case of product innovation, these numbers imply that the price increases very mildly over time at 1.4% per annum, while output is growing at 5.7 percent per year. Labour inputs to production increase at a rate of 7.1% per year, the same rate at which revenue is increasing. For process innovation, these same numbers yield an altogether different pattern. Price decrease sharply at 9.5% per year and output increases at 11.9% per year. Note that most of

this increase in output is owed to a more efficient production, since the labour force only increases by 2.4% per year. Revenue increases at this 'low' rate too. Although revenue seems to have a steeper trend for product innovation, we can't tell whether the rate of increase in profits is greater in this case, since this depends on the rate of growth in the costs of production and research. With respect to the wage bill in manufacturing, it is clear that labour costs are growing more rapidly in the case of product innovation (7.1 vs. 2.4% per year) but, lacking a figure for the rate of change of research workers, it is impossible to make a conclusive comparison. At any rate, the previous observations suggest that the financial picture of a firm, as revealed by its balance books, could be markedly different when it is directing its R&D towards product or process innovations.

3.6 Conclusions

This chapter has set up a dynamic model of exploratory- and applications-oriented research allocation by a monopoly enterprise as a constrained optimal control problem. Exploratory research, broadly conceived of as the search for an understanding of general phenomena, and the intangible knowledge that can be embodied in research platforms or research tools, is treated as contributing to the firm's profits only indirectly. The findings do not generate an immediate flow of revenue, but instead raise the productivity of applications-oriented researchers that are employed to develop product designs of better quality or more efficient production technologies A comparison of the results for a model in which the firm's exploratory research is complementary to new product-oriented R&D, or to 'new' process R&D, reveals that the central difference between the two is reflected in the optimised production plan. In either case, the optimal research plan implies that the firm will gain fundamental knowledge resources at a slowing pace in any specific research domain. Because the functional forms are very general and the assumptions they satisfy have been kept at a

minimum, the optimised evolutions followed by some variables are not completely pinned down. Nevertheless, it is possible to give a good indication of the upper bounds to their growth rate, and of the precise relationship that must hold between the rates of change of different variables.

The results emerging from the comparative examination of product and process innovations suggest that a profit-maximising firm having a monopoly position in its product market may go through a product or process "life cycle," albeit one with different underpinnings to the product life cycle theory originally proposed by William Abernathy and James Utterback, or that advocated more recently by Steven Klepper. In Abernathy and Utterback (1978), product innovation decreases as the industry evolves because of diminishing returns to investment in product innovations and lock-in by consumers to a "dominant design" for the product. Alternatively, Klepper (1996, p. 563) suggests that the decline in product innovation that occurs over time is a consequence of a "decline in the number of competitors brought about by the shakeout of producers" that follows from two assumptions: first, that the value of process innovations reducing the unit costs of production are proportional to the level of output of the firm; and second, that there are convex adjustment costs that give an advantage to early entrants. In these accounts, the technological life cycle is largely a result of competitive forces acting on the firms' incentives to enter the industry and innovate over time.

The model developed in this chapter also can accommodate the familiar S-shaped curve emerging from the product life-cycle theory, but the rationale for this profile lies in the nature of the complementarity between exploratory- and applications-oriented research activities. Specifically, an S-shaped curve for either the product quality design or the state of the production technology can result if the trajectory of applications-oriented research is single peaked; i.e., constant or increasing at the start of the firm's planning horizon and then

reversing course later on. To understand why this is the case one needs to consider results (R17) and (R17*), for product and process innovation respectively:

(R17) $\dot{N}_A \geq 0$ is sufficient for $\ddot{Z} > 0$

$\dot{N}_A < 0$ is necessary for $\ddot{Z} \leq 0$

$\ddot{Z} \leq 0$ is sufficient for $\dot{N}_A < 0$

(R17*) $\dot{N}_A \geq 0$ is sufficient for $\ddot{T} > 0$

$\dot{N}_A < 0$ is necessary for $\ddot{T} \leq 0$

$\ddot{T} \leq 0$ is sufficient for $\dot{N}_A < 0$

Result (R17) shows that improvements in product quality will be accelerating at the beginning of planning if applications-oriented research is constant or increasing and that it is necessary although not sufficient for this mode of research to be decreasing for product quality to grow at a decreasing rate. The same conclusion can be extracted from (R17*) for innovations in the production system design. But what is most interesting is that the optimised evolution of the state of fundamental knowledge – increasing rapidly at first and then slowing down, since $\dot{S} > 0$ and $\ddot{S} < 0$ – seems to push the firm exactly in the direction of adopting a single-peaked path for applications-oriented research. Recall that the optimal evolution of applications-oriented research is governed by the following equation in the case of product innovation:

$$\dot{N}_A = \frac{-1}{h_{11}(N_A, S)}\left[h_{12}(N_A, S)\dot{S} - \frac{p_1(Z,q)q h_1(N_A, S)^2}{w_A} \right]$$

The preceding equation suggests that product innovation-oriented research is more likely to be increasing at the outset, when knowledge resources are increasing more rapidly, because the marginal productivity of this mode of research will then be increasing at a faster pace too. As expenditures on exploratory research are 'phased out,' the state of knowledge advances more slowly, reducing the incentives for applications-oriented research searching for

improved product designs. The same arguments hold for the case of process innovation, as can be verified by the corresponding equation for applications-oriented research:

$$\dot{N}_A = \frac{-1}{h_{11}(N_A,S)}\left[h_{12}(N_A,S)\dot{S} - \frac{F(L)[p_2(1,q)q + p(1,q)]h_1(N_A,S)^2}{w_A}\right]$$

There is a completely different interpretation to the problem solved in this chapter that is interesting in its own right and, in addition, sets into sharp relief the first of several limitations of our set-up. The point of departure for this interpretation is that profit maximisation by an enterprise with a product-monopoly approximately corresponds to the problem faced by a central planner maximising a social welfare function under certain cost and demand conditions. This result is due to Arnold Harberger's 1954 analysis of monopoly and resource allocation. Harberger showed that the misallocation of resources by a monopoly is minimal in the absence of fixed costs (i.e., a setting where the marginal cost curve of the firm coincides with the average cost curve) when the demand for the product has unit elasticity. To the extent that one admits the reasoning Harberger puts forward, one also can accept the reading of the model's solution as a close approximation of the outcome from the maximisation of social welfare.[14]

Seen from this angle, the solution to our model may be used to make some comments on the path of exploratory and applications-oriented research activities decided upon by a social planner at the macroeconomic level. The natural question is whether the optimised

[14] This is because the assumptions made in this chapter about costs and demand can encompass Harberger's. This can be observed from the specification of the previous section. Strict concavity of the revenue function required $\varepsilon > 1$, but this can be relaxed to concavity $\varepsilon \geq 1$ without any significant changes to the results derived, and is still sufficient to guarantee a global maximum. Notice that Cowling and Mueller (1978) criticised Harberger for adopting these extreme assumptions. However, their criticisms are directed at the use of this description in an *empirical* analysis of social welfare losses provoked by market power for the U.S. economy, not at the *theoretical* validity of the argument that under certain cost and demand conditions welfare losses are necessarily small. Therefore, they do not represent a challenge to the correspondence between the problem of maximizing profits and that of maximizing social welfare.

trajectories for the control and state variables we have found seem to be sensible descriptions at the aggregate level. A technological life cycle may be a plausible description of research activities conducted by a firm developing a family of successively better products or processes, but is it a reasonable depiction of research investments in the whole economy? In answer to these questions, one can observe that evidence from the U.S. and other developed countries shows R&D expenditures to be increasing over the course of the 20th century, not decreasing.[15] As David, Hall and Toole (2000, pp. 497-8) have remarked: "the fraction of real gross national product being directed by both private and governmental agencies toward expanding the base of scientific and technological knowledge for non-defence purposes has trended upwards – halting at times, yet never reversing significantly."

This inconsistency between the theoretical results and the empirical evidence suggests that there are some unmodelled features of the innovation process at the national level that sustain the rate of return to exploratory research over time. Among these one could include a variety of learning processes, such as learning as a by-product of production, or improved educational courses and screening by employers, which improve the quality of research workers and their productivity over time. A related possibility is that, while this chapter looked at the exploitation of one product line, a firm or an aggregate of firms will develop *many* product-lines simultaneously or sequentially. There may be decreasing returns to the development of individual product lines, but a steady state in exploratory research could be justified if firms are switching across different lines of investment as time passes, particularly when there are knowledge spillovers or demand-side externalities across them. And the same

[15] Note in this regard that, whereas the conditions on the 'production functions' for fundamental knowledge and applications needed to ensure the existence of an interior solution were argued to be untenable at the firm-level, as it is evident that many firms do not invest in innovation, they acquire relevance if we consider the solution in the light of a social welfare maximization. At the industry or national level, the evidence points to positive amounts of exploratory and applications-oriented research being financed.

may be true of an economy taken as a whole if new industries displace mature ones where growth has petered out, as Simon Kuznets suggested many years ago:

> "As we observe various industries within a given national economy, we see that the lead in development shifts from one branch to another. A rapidly developing industry does not retain its vigorous growth forever but slackens and is overtaken by others whose period of rapid development is beginning. Within one country we can observe a succession of different branches of activity in the vanguard of the country's economic development, and within each industry we can notice a conspicuous slackening in the rate of increase."[16]

But tackling those issues without fully understanding the implications of the dynamic problem that has been put forward risks 'casting the net too widely.' Therefore, we will put on hold the analysis of R&D activities in settings with multiple product lines until Chapter 5., and try to address the more immediate limitations of the present analysis. It is evident that the best way to get the analysis around the limits imposed by a purely analytical framework and solution methods and generate a sharper picture of the firm's research and production decisions is by making a methodological switch. This is the course taken in Chapter 4, which embarks on a numerical investigation of a specified version of the model. By so doing, we will be able to validate the conjecture that a technological life cycle emerges as a result of a profit-maximising firm's optimal controls. The flexibility of the numerical tools will be useful in dealing with two other shortcomings of the study undertaken in this chapter.

First, the results derived so far are relevant to understanding the production schedule and research programme in an interior solution, but do not help us to discern the circumstances under which a firm finds exploratory research or applications-oriented research profitable. Moreover, assuming that the conditions for positive spending in both kinds of research are met means that the analysis has left part of the puzzle unsolved To understand the research programme in general would require knowledge of when and why the firm

[16] Kuznets (1954, p. 254). See also Kuznets (1966, 1971).

moves between 'boundary conditions,' where only one mode of research is supported, to the interior solution analysed here, to the case in which no research is financed at all. By examining the comparative dynamics of the optimal paths with numerical methods, it will be possible to provide some insights into the product market, firm level, and factor market conditions supporting a given pattern of 'research specialisation.'

Second, an attentive reader will have noticed that we have been working under the assumption that the firm does not discount its stream of profits. The reason is that discounting is not necessary to the solution of an optimal control problem with a finite planning horizon yet its inclusion introduces an additional complication, as the admissible interval for the discount factor would need to be conditioned to properly determine the properties of the trajectories followed by the costate variables (π_S, π_Z and π_T). Essentially, the discount needs to be 'small enough' so that the shadow prices are decreasing over time. As a first approximation to an already complicated problem, this may be an acceptable assumption and we are not alone in making it; Bayus (1995) uses a similar mathematical set-up to consider simultaneous product and process innovations, and derives his main proposition under the assumption that the discount rate is zero for the same reason. Still, discounting is an important aspect of any investment decision and there is reason to think it will be central to the assessment of R&D investment projects, particularly those involving an important element of exploratory research, as it typically takes years not months to see them come to full fruition. The manner in which a positive discount rate influences the scale of the firm's R&D spending, and the share of the resources supporting each mode of research, is another area that can be illuminated by the numerical methods employed in the following chapter.

Appendix. Notation Key

Market demand function

$p(Z(t),q(t))$ inverse demand function

$Z(t)$ product 'design' quality at time t

$q(t)$ quantity offered in the market at time t

$p_i(Z(t),q(t))$ derivative of the inverse demand function with respect to argument i, e.g.,

$p_{11}(Z(t),q(t))$ is the second derivative with respect to product quality

Production function

$q(t)=T(t)F(L(t))$ firm's production function

$T(t)$ efficiency of the production technology at time t

$L(t)$ labour force employed in production at time t

$T(t)F'(L(t))$ marginal product of labour

Product quality

$\dot{Z}(t)=h(N_A(t),S(t))$ evolution of product design quality

$N_A(t)$ labour employed in applications-oriented research at time t

$S(t)$ knowledge available to the firm at time t

$h_i(N_A(t),S(t))$ derivative of $h(N_A(t),S(t))$ with respect to argument i

$\pi_Z(t)$ shadow price of quality additions

Production system designs

$\dot{T}(t)=h(N_A(t),S(t))$ evolution of production system designs

$\pi_T(t)$ shadow price of additions to the state of production technology

Firm's knowledge resources

$\dot{S}(t) = g(N_E(t))$ growth of knowledge resources

$N_E(t)$ labour employed in exploratory research at time t

$\pi_S(t)$ shadow price of additional knowledge resources

Labour supply

$w > 0$ constant wage of production labour in competitive labour market

$w_A > 0$ constant wage of applications-oriented research labour in competitive labour market

$w_E > 0$ constant wage of exploratory research labour in competitive labour market

Extra notation for example

$p(Z(t), q(t)) = Z(t)^{\chi/\varepsilon} q(t)^{-1/\varepsilon}$ inverse demand function

$\varepsilon > 0$ elasticity of demand with respect to price

$\chi > 0$ elasticity of demand with respect to product quality

$q(t) = T(t)L(t)^{\alpha}$ firm's production function, α is constant with $0 < \alpha < 1$

$\dot{Z}(t) = N_A(t)^{\nu} S(t)^{1-\nu}$ evolution of product quality

$\dot{T} = N_A(t)^{\nu} S(t)^{1-\nu}$ evolution of production technology

ν is constant with $0 < \nu < 1$; ν is the share of research labour in applications-oriented research, be it for improved products or manufacturing processes, $1-\nu$ is the share of scientific knowledge in the creating improved product or production system designs

$\dot{S}(t) = N_E(t)^{\phi}$ growth of knowledge resources, ϕ is constant with $0 < \phi < 1$

Growth rate of variable i indicated by γ_i

Chapter 4

Firm-level R&D Expenditures under Monopoly Conditions: A Numerical Analysis

4.1 Introduction

The previous chapter set up a dynamic model of exploratory- and applications-oriented research allocation by a monopoly enterprise as a constrained optimal control problem. The assumption of a product monopoly put aside strategic and other considerations in order to focus attention on the way a firm structures its R&D programme over time, given the dynamic connections between the heterogeneous research activities and production. Exploratory research, broadly conceived of as a search for an understanding of general phenomena, and the intangible knowledge that can be embodied in research platforms or research tools, is treated as contributing to the firm's profits only indirectly. Its findings do not attract an immediate flow of revenue, but instead may be used by applications-oriented researchers labouring to improve product designs or the production technology. A comparison of the results for a model in which the firm's exploratory research is complementary to new product-oriented R&D, or to 'new' process R&D, reveals that the central difference between the two is reflected in the optimised production plan. The optimal research plan implies that the firm will gain fundamental knowledge resources at a slowing pace. Practical knowledge about product quality or process efficiency could have an S-shaped evolution, consistent with a technological life-cycle interpretation.

With very general production functions for goods and knowledge, it was impractical to aim for an explicit solution, since the various ties between the co-state, state and control variables generated elaborate dynamic movements in the solution. To deal with this, the

optimised trajectories were depicted within the bounds marked by several assumptions, in particular that an interior solution existed and that the profit stream was not discounted. A numerical study of the problem using a particular specification would be justified simply to get a sharper picture of the firm's research and output decisions across a "family" of improving products or one of manufacturing processes. In addition, a numerical analysis is the ideal setting in which to investigate the influence of product and factor market conditions, firm-level characteristics, and discounting on the optimal research and production programme. It will be possible to go beyond the standard comparative analysis of equilibrium dynamics to investigate the re-optimisation of the development trajectories when the firm experiences shocks affecting parameters or state variables after it has disbursed part of the R&D investments originally calculated to maximise profits.

The order and key results of the chapter are as follows. Section 4.2 begins by recapitulating the optimal control problem representing the firm's dynamic decisions. Section 4.3 spells out the specific functions and the benchmark parameter values used in the numerical analysis, and gives an expeditious account of the numerical methodology. The analysis rests on the use of a boundary value problem algorithm to solve the system of ordinary differential equations that are derived from the necessary conditions for a global maximum. The fourth section describes the general features of the numerical solution to the problem. The solution is characterised by an initially high but declining rate of investment in exploratory research, and a subsequently rising and then declining rate of application-oriented R&D. Therefore, by the end of the project planning interval the firm is completely dedicated to manufacturing with no further R&D spending.

Section 4.5 considers the results from a comparative dynamic analysis. This indicates that a firm that faces high interest rates or a small market for its product, is very inefficient in its production or research, or must pay high wages to researchers is more likely to 'cancel' its

exploratory research. When these constraining conditions are very pronounced, the firm also may dispense with commercially oriented research on applications. Section 4.6 extends the numerical analysis to study the allocation of R&D in a volatile environment, where the firm experiences exogenous shocks affecting market conditions or its internal research capabilities midway through the course of a cycle of development. As the various classes of knowledge are inherited from the past, the revised R&D and production plans exhibit hysteresis: their 'continuation' paths differ from those that would be found had the trajectory of the system been optimised subject to the different parameter values from the start. Section 4.7 concludes with a summary of the results, a discussion of the implications of this analysis and ways in which it could be extended.

4.2 Outline of the dynamic model

Given the intricacies of the model of exploratory- and applications-oriented R&D, and that this chapter explicitly considers the effects of discounting, it is advisable to go over the model's structure and the necessary conditions for its solution in brief terms. For an unabridged description of the model's notation and assumptions (in the case of profit-maximisation without discounting), see Sections 3.2, 3.3, and 3.4 in Chapter 3.

A firm seeks to determine the research and production plan that maximises its cumulative, discounted profits over a fixed planning horizon $[t_0, t_1]$. The exponential discount factor is $e^{-\delta t}$, where δ is the discount rate. The firm is in possession of a 'product design' for a good of quality $Z(t_0)$, which it alone can manufacture by employing manufacturing labour $L(t)$, conditional on the initial production technology state $T(t_0)$. Output at time t is determined by the production function $q(t) = T(t)F(L(t))$, where $F(L(t))$ is increasing and concave in $L(t)$. The market demand faced by the firm is represented by an inverse demand

function $p(t) = p(Z(t), q(t))$; it is increasing in the product's quality $Z(t)$ and decreasing in the quantity supplied $q(t)$.

The firm can hire research workers to engage in either product innovation oriented research or (own)-process oriented research, both of which are modes of *applications-oriented research*. Employing applications-oriented researchers $N_A(t)$ generates increases in the quality of the product $\dot{Z}(t) = h_Z(N_A(t), S(t))$, or the efficiency of the production technology $\dot{T}(t) = h_T(N_A(t), S(t))$ that are conditional on the state of knowledge $S(t)$. The function $h(N_A(t), S(t))$ is increasing in both arguments and jointly concave in $N_A(t)$ and $S(t)$; researchers and knowledge are complements in innovation, $h_{12}(N_A(t), S(t)) > 0$. The firm can utilise the innovations generated by the research it funds without the need to pay for any adjustment costs. Notice that, while an improved product design raises the firm's revenue thanks to an increase in demand for the commodity it manufactures, a superior efficiency of its process technology boosts its production of final goods for a given number of workers.

At the same time, the firm can invest in *exploratory research* $N_E(t)$ to supplement the fundamental knowledge $S(t)$ available to, and hence the productivity of, research for 'applications.' The evolution of the state of fundamental knowledge is described by $\dot{S}(t) = g(N_E(t))$, where the function $g(N_E(t))$ is increasing and concave in $N_E(t)$. At the outset, the state of knowledge is $S(t_0)$ reflecting the research workers tacit knowledge, as well as the collection of knowledge resources that are publicly available. As with technology, the firm can make use of any fundamental knowledge it obtains without further outlays.

Summarising the problem, the objective of the firm is to select time paths for its controls $L(t)$, $N_A(t)$, and $N_E(t)$ to maximise the discounted sum of profits from the initial time t_0 to the terminal time t_1:

$$\Pi = \int_{t_0}^{t_1} e^{-\delta t}\left[p(Z(t),q(t))q(t) - wL(t) - w_A N_A(t) - w_E N_E(t)\right]dt$$

where $q(t) = T(t)F(L(t))$. The maximisation is subject to:

1) either $\dot{Z}(t) = h_Z(N_A(t), S(t))$ or $\dot{T}(t) = h_T(N_A(t), S(t))$, depending on whether the firm means to improve its product quality or manufacturing efficiency respectively,

2) $\dot{S}(t) = g(N_E(t))$, the equation of motion describing the relationship between the state of knowledge and exploratory research employment,

3) nonnegativity constraints on the three control variables $L(t) \geq 0$, $N_A(t) \geq 0$ and $N_E(t) \geq 0$,

4) and boundary conditions on the state variables, namely the initial value for product quality $Z(t_0) = Z_0 > 0$, the production technology state $T(t_0) = T_0 > 0$, and the fundamental knowledge state $S(t_0) = S_0 > 0$, leaving the terminal values $Z(t_1)$, $T(t_1)$ and $S(t_1)$ free.

The model is therefore a constrained optimal control problem with three control and two state variables. Two current-value Hamiltonians can be set up, one for the product innovation case and the other for process innovation, and the usual necessary conditions will indicate the solution. To illustrate the kind of conditions this imposes, consider the case of an improving product family, for which the current-value Hamiltonian function is:

$$H_C = p(Z(t), T(t_0)F(L(t)))T(t_0)F(L(t)) - wL(t) - w_A N_A(t) - w_E N_E(t) \qquad (4.1)$$
$$+\psi_Z(t)h(N_A(t), S(t)) + \psi_S(t)g(N_E(t))$$

where $\psi_Z(t)$ is the current-value costate variable corresponding to the differential equation of motion $\dot{Z}(t)$ and, analogously, $\psi_S(t)$ is the current-value costate variable associated to $\dot{S}(t)$. In this case, the state of the production technology is the same throughout the planning horizon, $T(t) = T(t_0)$. One can then define a Lagrangian function,

$$\mathsf{L} = H_C + \lambda_L(t)L(t) + \lambda_{N_A}(t)N_A(t) + \lambda_{N_E}(t)N_E(t)$$

$$\begin{aligned}\mathsf{L} = {} & p(Z(t), T(t_0)F(L(t)))T(t_0)F(L(t)) - wL(t) - w_A N_A(t) - w_E N_E(t) \\ & + \psi_Z(t)h(N_A(t), S(t)) + \psi_S(t)g(N_E(t)) \\ & + \lambda_L(t)L(t) + \lambda_{N_A}(t)N_A(t) + \lambda_{N_E}(t)N_E(t)\end{aligned} \tag{4.2}$$

where $\lambda_L(t)$, $\lambda_{N_A}(t)$ and $\lambda_{N_E}(t)$ are Lagrange multipliers corresponding to the three nonnegativity constraints on $L(t)$, $N_A(t)$ and $N_E(t)$, respectively. The optimal solution must satisfy the following necessary conditions (so as to simplify the notation we get rid of the (t)'s and suppress the asterisk for the control, state and costate variables, but bear in mind that they refer to the optimal paths):

$$\frac{\partial \mathsf{L}^*}{\partial L} = p_2(Z, T(t_0)F(L))T(t_0)^2 F'(L)F(L) + p(Z, T(t_0)F(L))T(t_0)F'(L) - w + \lambda_L = 0 \tag{4.3}$$

$$\frac{\partial \mathsf{L}^*}{\partial N_A} = -w_A + \psi_Z h_1(N_A, S) + \lambda_{N_A} = 0 \tag{4.4}$$

$$\frac{\partial \mathsf{L}^*}{\partial N_E} = -w_E + \psi_S g'(N_E) + \lambda_{N_E} = 0 \tag{4.5}$$

$$\begin{aligned}&\lambda_L \geq 0 \qquad L \geq 0 \qquad \lambda_L L = 0 \\ &\lambda_{N_A} \geq 0 \qquad N_A \geq 0 \qquad \lambda_{N_A} N_A = 0 \\ &\lambda_{N_E} \geq 0 \qquad N_E \geq 0 \qquad \lambda_{N_E} N_E = 0\end{aligned} \tag{4.6}$$

$$\dot{\psi}_Z = -\frac{\partial \mathsf{L}^*}{\partial Z} + \delta\psi_Z = -p_1(Z, T(t_0)F(L))T(t_0)F(L) + \delta\psi_Z \tag{4.7}$$

$$\dot{\psi}_S = -\frac{\partial \mathsf{L}^*}{\partial S} + \delta\psi_S = -\psi_Z h_2(N_A, S) + \delta\psi_S \tag{4.8}$$

$$\dot{Z} = \frac{\partial \mathsf{L}^*}{\partial \psi_Z} = h(N_A, S) \tag{4.9}$$

$$\dot{S} = \frac{\partial \mathsf{L}^*}{\partial \psi_S} = g(N_E) \tag{4.10}$$

$$Z(t_0) = Z_0 > 0 \qquad e^{-\delta t_1}\psi_Z(t_1) = 0 \tag{4.11}$$

$$S(t_0) = S_0 > 0 \qquad e^{-\delta t_1}\psi_S(t_1) = 0 \tag{4.12}$$

As Chapter 3 discussed at some length, the existence of a solution can be guaranteed by making additional assumptions about the concavity of the objective function and the equations of motion. The following section gives more details of the precise requirements in a specified version of the model.

4.3 Numerical methodology

4.3.1 Benchmark specification

To use the computational tools available for this kind of control problem, it is necessary to define specific forms for the functional relations postulated in section 4.2. We will employ the functional forms introduced in Section 3.5 of the preceding chapter:

(1) The inverse demand function is specified as $p(Z,q) = Z^{\chi/\varepsilon} q^{-1/\varepsilon}$, where $\varepsilon > 0$ is the elasticity of demand with respect to price and $\chi > 0$ is the elasticity of demand with respect to product quality, as can readily be observed from the corresponding Marshallian demand function $q(Z,p) = Z^{\chi} p^{-\varepsilon}$.

(2) The specification for the firm's production function is $q = TL^{\alpha}$, where $0 < \alpha \leq 1$.

(3) The evolution of product quality is governed by $\dot{Z} = N_A^{\nu_Z} S^{1-\nu_Z}$, with $0 < \nu_Z < 1$; where ν_Z is the share of applications-oriented research labour and $1-\nu_Z$ is the share of scientific knowledge in advancing the product design.

(4) Production technology evolves, similarly, according to $\dot{T} = N_A^{\nu_T} S^{1-\nu_T}$, with $0 < \nu_T < 1$; ν_T is the share of applications-oriented research labour and $1-\nu_T$ is the share of scientific knowledge in advancing the production system designs.

(5) The transition equation for the state of knowledge is $\dot{S} = N_E^{\phi}$, where $0 < \phi < 1$.

The iso-elastic functional forms are frequently used in numerical analysis because of their flexibility and the simple interpretation of the parameters as elasticities, which can be helpful in calibrating the model to data. In order to satisfy the sufficient conditions for the existence of a solution, additional restrictions must be imposed on the values taken by the parameters. Specifically, the revenue function needs to be jointly concave in each of its arguments, and we shall assume it is strictly concave. In the case of product innovation, the conditions for strict concavity of the revenue function are $\varepsilon > \chi$, $\varepsilon > 1$, and $\alpha < [\varepsilon - \chi]/[\varepsilon - 1]$. The upper bound on α is redundant if $0 < \chi < 1$, as this parameter needs to be constrained to $0 < \alpha \leq 1$ anyhow, but binding whenever $\chi \geq 1$. When we turn to improvements in the production system, the relevant conditions are $\varepsilon > 1$ and $\alpha < 1/[\varepsilon - 1]$; if the price elasticity of demand is in the open interval $1 < \varepsilon < 2$, the latter condition is redundant. [1]

These functional forms need to be substituted into the necessary conditions listed above. The resulting conditions can then be simplified to get a reduced number of differential equations that are functions of the state variables (S and Z, or T depending on the type of

[1] The reader is referred to Sections 3.3 and 3.4 in Chapter 3 for a detailed description of the sufficient conditions for existence in the general formulation, and to Section 3.5 for a discussion of their economic meaning in the case of the specific functions employed in these simulations.

innovation) and costate variables (ψ_S and ψ_Z, or ψ_T) alone. Appendix 4.1 presents the reduced systems of differential equations corresponding to product and process innovation. One also needs to specify the boundary conditions at the outset and end of the planning horizon. The system of non-linear ordinary differential equations (ODE's) and boundary conditions make up a two-point boundary value problem. The optimal paths for the state and costate variables can be solved with Matlab; Section 4.3.2 below gives some details of Matlab's boundary value solver. From these paths it is possible to calculate the optimised values for the control variables, revenue, R&D expenditures and profits.

Table 4.1 Functional forms

1. *Inverse demand function* $p(Z(t),q(t)) = [Z(t)]^{\chi/\varepsilon}[q(t)]^{-1/\varepsilon}$ $\varepsilon > 0$ price elasticity of demand $\chi > 0$ elasticity of demand with respect to product quality
2. *Firm's production function* $q(t) = T(t)[L(t)]^{\alpha}$, $0 < \alpha \leq 1$
3. *Evolution of product quality* $\dot{Z}(t) = [N_A(t)]^{\nu_Z}[S(t)]^{1-\nu_Z}$, $0 < \nu_Z < 1$
4. *Evolution of production technology* $\dot{T}(t) = [N_A(t)]^{\nu_T}[S(t)]^{1-\nu_T}$, $0 < \nu_T < 1$
5. *Growth of knowledge resources* $\dot{S} = N_E^{\phi}$, $0 < \phi < 1$

One must then go through a trial-and-error process to find parameter values that yield a solution that is both 'stable' and economically interesting. The parameter values finally selected for the 'benchmark specification' are listed in Table 4.2. It is worthwhile making some observations about these parameters. First, as there is no information about the values taken by the elasticities determining the evolution of product quality and process efficiency, the simulations are run under the restriction that $\upsilon_Z = \upsilon_T$. Second, to ease comparison between the optimal trajectories for product and process innovation, the benchmark parameters only differ in the value taken by the price elasticity of demand; by setting the

demand elasticity slightly lower for product innovation, the optimised applications research spending has an interior peak in both simulations. Third, there is a wage gap between workers employed in manufacturing and those in research, but not between researchers engaged in exploratory and applications research. Fourth, the discount rate is set to 0 for the benchmark case. Note that each of these restrictions will be relaxed in the section examining the comparative dynamics, which investigates the problem with positive discount rates. Lastly, the numerical integration is executed across 25 time periods; i.e., $t_1 = 25$.

Table 4.2 Parameter values for 'benchmark specifications'

Parameter	ε	χ	α	υ_Z	υ_T	ϕ	w	w_A	w_E	δ	$Z(0)$	$T(0)$	$S(0)$
Product	1.25	.75	.75	.5	–	.75	1	2	2	0	100	100	100
Process	1.55	.75	.75	–	.5	.75	1	2	2	0	100	100	100

4.3.2 Boundary value solver

The solution to the differential equations was achieved using Matlab's solver *bvp4c*, which is tailor-made to tackle two-point value boundary value problems (BVP's), including those that are non-linear and possibly 'stiff'[2]. The problem under consideration is non-linear but not 'stiff.' This code requires an initial guess to the solution, in addition to the system of differential equations and the boundary conditions. In the dynamic problem at hand this was easily specified, as there was a clear indication of the sign of the first and second derivatives of the state and co-state variables in the optimal solution. The reason for making such a guess is that BVP's, unlike initial value problems, can have multiple solutions. For the technically

[2] 'Stiff' problems are those in which the solved trajectories have abrupt if not discontinuous changes at some point or points in time. More formally, Guderley (1975, p. 417) defines a stiff two-point boundary value problems as that in which "the homogeneous system pertaining to the variational problem possesses particular solutions which decrease quickly as the independent variable increases and others which decrease quickly in the opposite direction."

minded, Appendix 4.2 provides, as an example, the Matlab program that was run to derive the results for the specification with product innovation. The way in which Matlab's algorithm works is described in simple terms in their online documentation as:

> The collocation technique uses a mesh of points to divide the interval of integration into subintervals. The solver determines a numerical solution by solving a global system of algebraic equations resulting from the boundary conditions, and the collocation conditions imposed on all the subintervals. The solver then estimates the error of the numerical solution on each subinterval. If the solution does not satisfy the tolerance criteria, the solver adapts the mesh and repeats the process.[3]

More technical details of *bvp4c* are beyond the scope of this section,[4] but notice that other methods are available to solve BVP's. Many programs implement "shooting codes," which first set up an initial value problem and then iteratively adjust the starting values until the solution is such that the desired boundary conditions are approximately satisfied.

4.3.3 Sensitivity analysis

There are several elements to the sensitivity analysis undertaken to verify the stability of our simulations. The first involved setting up an initial value problem with the same differential equations that appear in Appendix 4.1, into which were substituted the initial values for the two co-state variables revealed by Matlab's collocation method. This problem was solved via Mathematica's shooting code *NDSolve*, which alternatively uses the Adams or the Gear method depending on the 'stiffness' of the problem. The results for the trajectories of the state and co-state variables from this problem were identical to those found by Matlab's algorithm and the terminal boundary conditions were satisfied, substantiating the accuracy of that numerical solution. This test does not prove that this is the only solution, but it is

[3] See http://www.mathworks.com/access/helpdesk/help/techdoc/math_anal/diffeq13.shtml for the documentation.

[4] For a clear description of Matlab's solver and its applications see Shampine, Kierzenka and Reichelt (2000). Rigorous mathematical examinations of the Lobatto collocation method exist in, e.g., Russell and Christiansen (1978) or Ascher and Weiss (1983, 1984), but they are not for the faint-hearted.

encouraging that the two approaches to the numerical analysis of non-linear two-point boundary value problems produce identical results.

Another, more informal sensitivity test was to evaluate the solution in the light of the analytical results of the previous chapter. Without going into the details of this comparison, as these are described in the following section, the numerical solutions in the benchmark specifications revealed trajectories for the costate, state and control variables that satisfy all of the properties derived in the formal analysis of the problem. And there is little risk that using the selfsame analytical results as the basis for the initial guess structure influenced the initial mesh constructed by Matlab to such an extent that the end results were biased too. Careful testing showed that the solution found using *bvp4c* was robust to different guess structures. The numerical solutions to the problem share the same *properties* as the initial guess (e.g., the state of knowledge is in both cases increasing) but by no means were the numbers supplied and those obtained close to each other. For example, the guess was simply that the state of knowledge would grow from a value of 100 at $t_0 = 0$ to 125 by the terminal time $t_1 = 25$, while in the benchmark specification it went all the way to 300 and under different parameter values the end figure could be much greater.

The comparative dynamic analysis provided a check of the stability of the results to parameter changes. Although numerical analyses of non-linear differential equations are generally sensitive to the choice of parameters, a well-formulated problem should at the very least be locally stable and the greater the range of parameters it can support without breaking down the more trustworthy it is. The model performed well in the examination of 'local stability,' but there were two separate cases in which we were unable to get meaningful results for larger variations in the parameters. First, the solution was found to be quite sensitive to the values of ϕ, υ_Z and υ_T, the parameters in the 'production functions' for fundamental knowledge and either product quality or production efficiency, and occasionally

generated complex-numbered solutions for values different to the 'benchmark.' If one goes back to the original differential equations (listed in Appendix 4.1) to understand how the two parameters influence the solution, an explanation can be given for these results. Consider the equation describing the evolution of the state of knowledge, which is the same for both product and process innovation:

$$\dot{S} = \left[\frac{\phi \psi_S}{w_E} \right]^{\frac{\phi}{1-\phi}}$$

With the benchmark value for $\phi = 0.75$ there is a cubic relationship between the evolution of the state of knowledge in equilibrium and the shadow price of knowledge ψ_S, and the solver produces real-valued results, as it does for small variations around this value. What is more, these can be given a natural economic interpretation, as we argue in Section 4.5. But the problem is more challenging for values further away from $\phi = 0.75$ because the system becomes more nonlinear, and imaginary numbers are commonly found as part of the solution. To informally test the validity of this argument, several simulations were run with values of ϕ for which the power to which ψ_S is raised is an integer, which include the fractions in the series: 1/2, 2/3, 3/4, 4/5, ... All of these simulations resulted in real numbered results. When ϕ takes a value that is very different to these fractions, Matlab's solver is forced to find a maximum to a problem that is no longer linear, quadratic or cubic but somewhere in between. The same reasoning and findings are true of υ_Z and υ_T, parameters that are fundamental to the equilibrium evolution of the costate and state variables.[5]

[5] This line of reasoning gains some force if one surmises that α, ε and χ do not have this 'perverse' influence because the sufficient conditions for the dynamic profit maximization restrict these parameters in such a way that the powers to which the variables are raised in the differential equations $\dot{\psi}_Z$ and $\dot{\psi}_T$ are always in the interval $(-1,0)$. Concave functions are usually easier to deal with.

Beyond the intrinsic difficulties of solving very non-linear systems of differential equations, notice that the complex-numbered solutions may be a sign that the firm's optimal controls are not monotonic under certain parameter values, but rather may follow a sine-like pattern, circling around a trend determined by the real part of the results. This interpretation is certainly not excluded by the sufficient conditions for the constrained optimal control problem: they guarantee that a solution will exist, but do not restrict the profile of the optimal trajectories followed by the state and control variables, which may 'oscillate.'

The second instance in which Matlab's algorithm did not produce well-behaved trajectories is less remarkable: while the solver can provide a the solution to the dynamic system's equilibrium evolution that satisfies the 0.1% relative tolerance levels for a broad range of the remaining parameters (χ, ε, α, w, w_A, w_E, $S(0)$, $Z(0)$, $T(0)$), there are values for which this cannot be met, simply because the state variables grow so rapidly. This is to be expected and doesn't pose a problem to the analysis in this chapter because the admitted range for the parameters is economically interesting and, at any rate, the problem could be re-scaled if one was interested in understanding the dynamic pattern when the elasticities reach 'extreme' values. As a related point, observe that the sizeable swings portrayed in the section examining the comparative dynamics are a direct result of the simplicity of the functional forms used in the numerical simulations. For a calibrated model, one would unquestionably include extra parameters to 'dampen' these changes.

4.4 Central results in the benchmark simulations

As mentioned above, the numerical solutions for the benchmark specifications, in which the discount rate is zero, reveal trajectories for the costate, state and control variables satisfying all the properties derived analytically. Recall among these general findings that exploratory research N_E was found to be positive and decreasing over time; applications-oriented

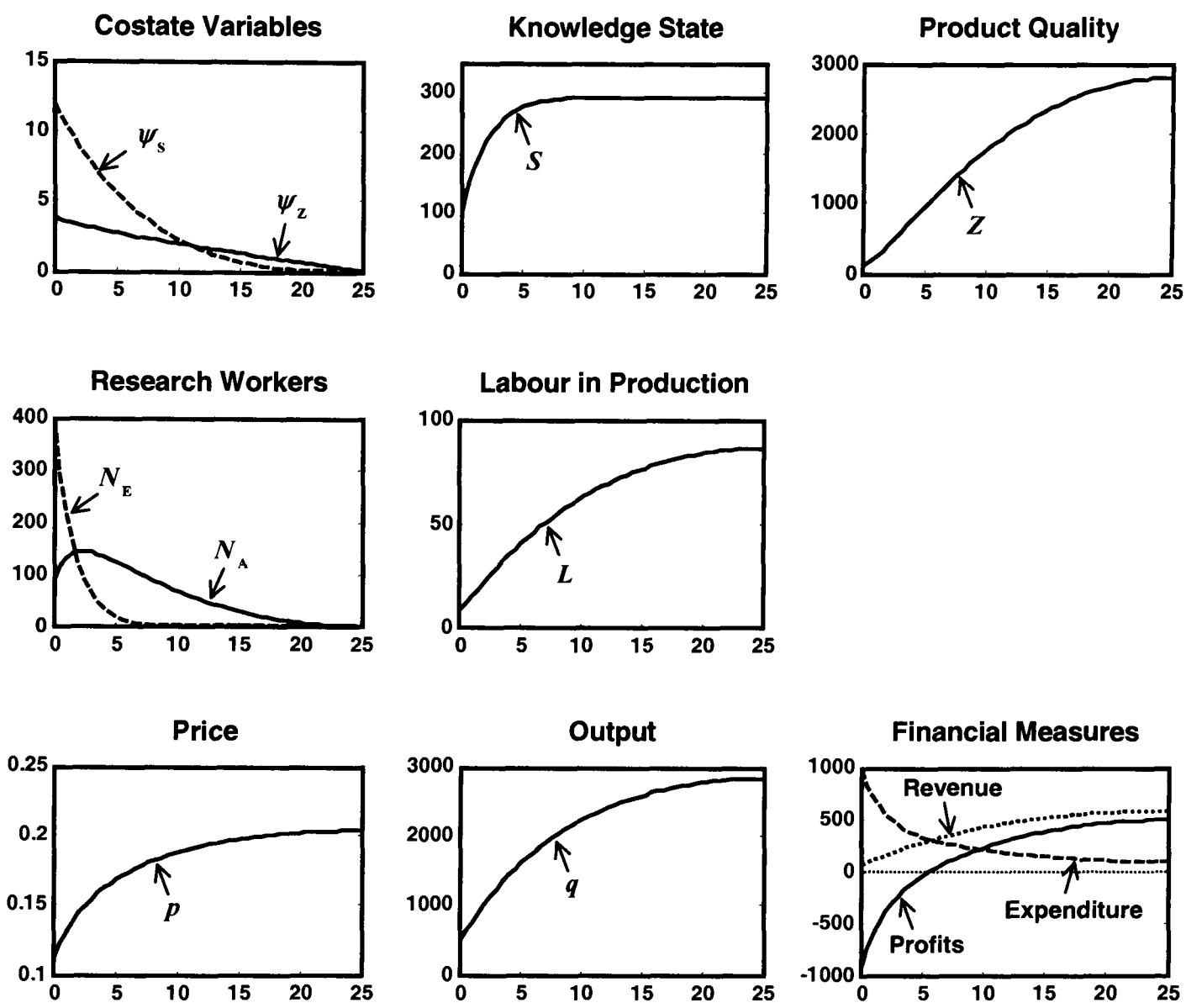
research N_A was positive but could be decreasing, constant or increasing. This path for the control variables implied that fundamental knowledge (S) and 'practical' knowledge embodied in applications (Z or T) were increasing; the former at a decreasing rate, the latter in a more complex manner that had as an interesting possibility an S-shaped evolution. The shadow value of fundamental knowledge additions, which was denoted π_S when there was no discounting, was proved to be positive and decreasing – and fell at a decreasing rate if the production function for applications $h(N_A, S)$ was homogenous of degree 1. The shadow value of quality improvements π_Z or process improvements π_T was in all cases positive and declined over time at a decreasing rate. All of these properties can be observed in the numerical solutions displayed in Figures 4.1 and 4.2.[6]

The interesting question, of course, relates to the new insights from having a sharper picture of the optimised decisions of the firm. In these simulations, the number of researchers allocated to exploratory research declines to approximately zero before the middle of the planning horizon. The evolution of expenditures on applications R&D is single-peaked, as conjectured in the previous chapter, with the peak occurring before exploratory research is exhausted. Applications-oriented research falls off to almost zero as the planning horizon comes to a close, but it can be increasing at the start of planning. To gain the most advantage from its store of fundamental knowledge, a firm would wait until exploratory research has pushed this knowledge frontier forward before searching more intensively for products of better quality or improved manufacturing processes. Not only can one tell that the firm is shifting resources from the exploratory mode of research to the commercially oriented one

[6] These figures also exhibit the specific results derived in Section 3.5, which analysed a specified model using the same functional forms as this chapter. This analysis showed that the main contrast between product and process innovations resided in the optimal decisions about the rate of change of prices: when the firm advanced the quality embodied in its products, output and prices increased concurrently; when the focus of this research was in the improvement of the manufacturing processes, output increased but the price set for the commodity declined.

that can attract an immediate flow of revenue, but also that cumulative R&D expenditures and the intensity of R&D effort (measured by the ratio of R&D spending to sales) decrease monotonically.

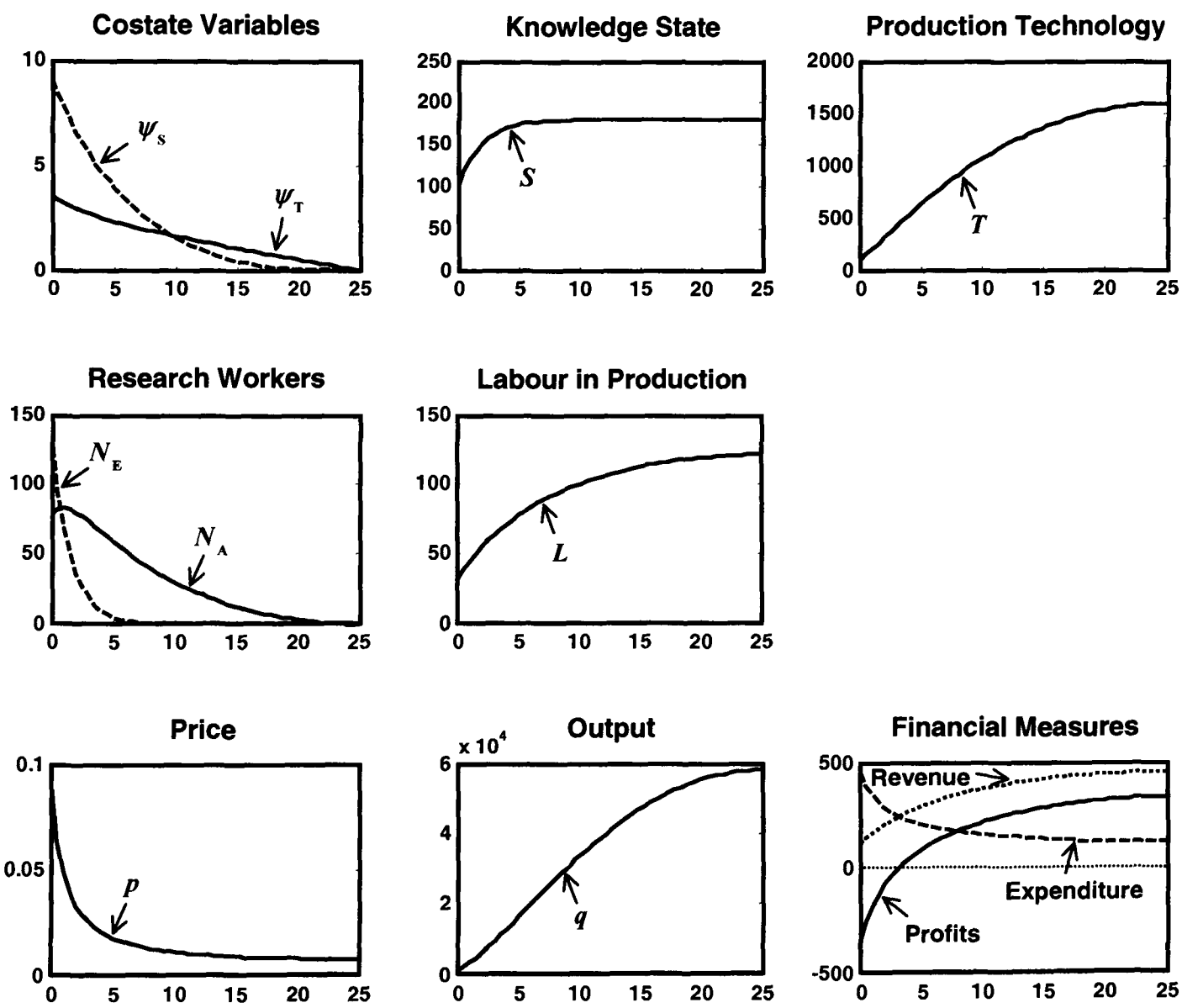
Figure 4.1 Numerical solution with product innovation



On the production side, the numerical results confirm that the price paid by consumers is decreasing when the firm is improving its production technologies and increasing when it manufactures products of increasing quality and, in addition, that the rate of change is very steep at the start of the planning horizon. The technical rationale for the sharp descent in the commodity's price is that, even with increasing labour productivity, the firm finds it optimal to hire more employees to work in manufacturing. A simultaneous increase in employment and labour productivity results in a rate of increase in sales exceeding that of manufacturing

efficiency. This explanation points to the close correspondence between the evolution in prices and output working through the market demand function. But the price trajectories exposed are of economic interest as well, as they seem to substantiate the assertion that "[t]he price of the product is one of the tools to push diffusion," as Perez and Soete (1988, p. 463) have put it. A different and perhaps more novel proposition can be put forward by moving away from the notion of diffusion, and moving towards that of private funding of innovation: price flexibility is a critical 'tool' for an innovating firm with a product monopoly to finance its exploratory- and applications-oriented research.

Figure 4.2 Numerical solution with process innovation

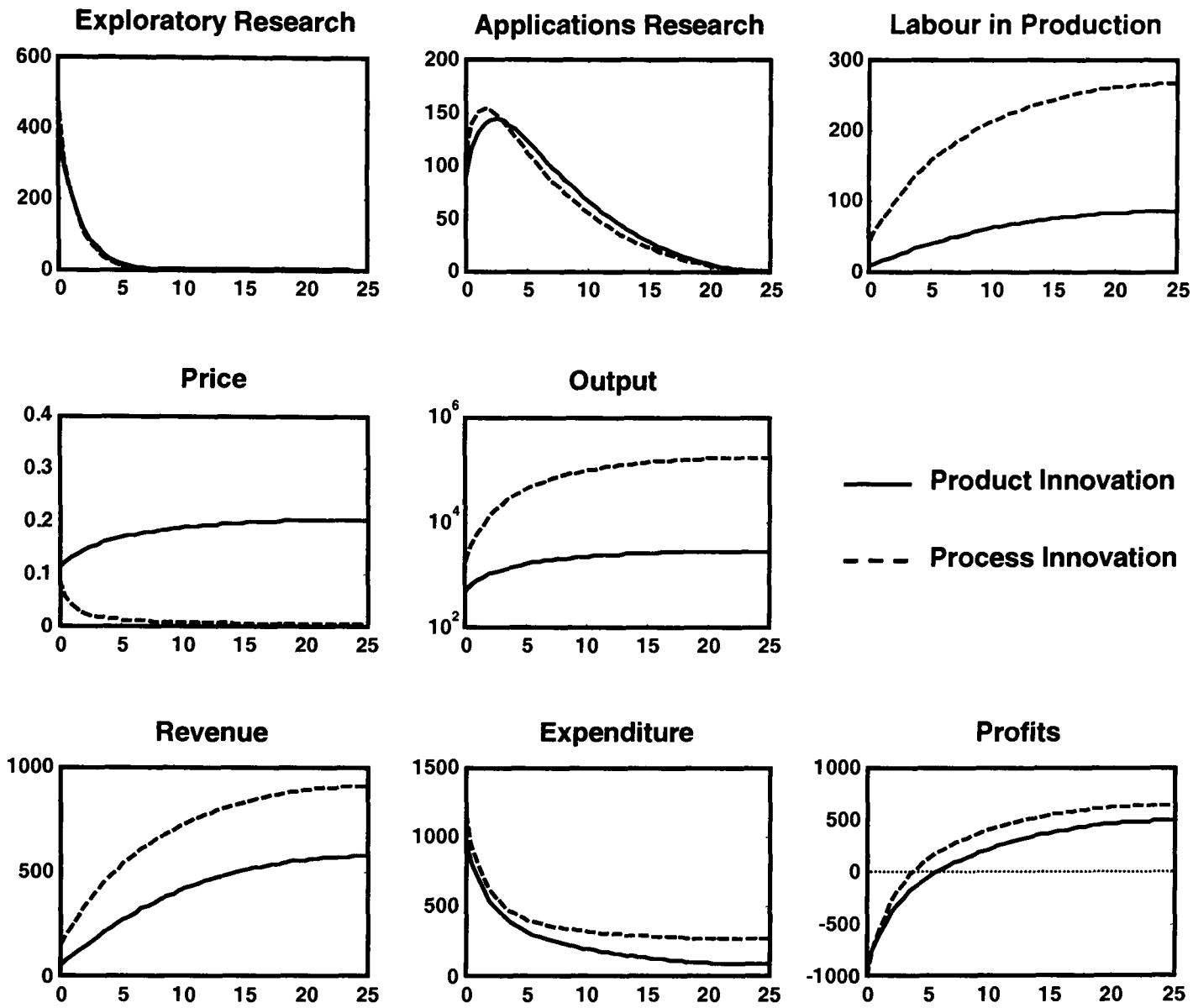


Economists confronted with this result might react with scepticism and ask: But shouldn't a monopolist set a constant (high) price in the case of *process* innovation? Does this not imply

that the firm is conducting an intertemporal variety of price discrimination, since the product's quality is the same throughout? In answer to the first question, notice that a monopolist improving its manufacturing processes may find it optimal to move along the demand curve as its costs of production are reduced, and this entails setting a lower price for the good it sells. In response to the second, the reason why consumers accept a higher price at the outset for an otherwise identical product is a consequence of adopting a time-invariant market demand function, which implicitly assumes that consumers are myopic and do not know the price will be cut later on. This unrealistic feature of the model is shared by many other dynamic formulations of innovation in the economic literature. It will be discussed at further length in the conclusions to this chapter, which presents alternative modelling assumptions that would not lead to this form of price discrimination.

Interestingly, although output and price follow different trajectories depending on whether the firm is undertaking product or process innovations, the firm's financial performance – in terms of its revenue, expenditure and profits – is remarkably similar in either case. Revenue is monotonically increasing and the total wage bill decreases monotonically. The upshot is that profits increase monotonically over time, but are negative at the start; this reflects the magnitude of the investments made early on by the firm and the increasing rewards gained afterwards, when the firm is predominantly concerned with manufacturing. The break-even point occurs early on, usually before the firm has exhausted its search for fundamental knowledge and not far apart from the peak in applications-oriented R&D. As we showed in Chapter 3, a growing revenue stream will always be observed in the solution with the specific functions we are using. The fall in labour expenditures and the position of the break-even point are reliant on the values of the parameters, and the comparative dynamics will give a clearer indication of their 'driving factors.'

Figure 4.3 Comparative examination of product and process innovation



To make the differences (and similarities) between the optimal programmes for process and product innovation more noticeable, one can run a simulation in which R&D expenditures across time are approximately the same for both kinds of innovation. One way of 'forcing' the schedules for exploratory- and applications-oriented research to roughly coincide is to adjust the value of price elasticity of demand in one of the two simulations. Setting $\varepsilon = 1.645$ in the simulation corresponding to process innovation arrives at the desired result, as can be seen in Figure 4.3, which depicts the complete numerical solution to this 'experiment.' The co-incidence of the two curves at this demand elasticity is not accidental: when $\varepsilon = \chi + 1$, the differential equation describing the evolution of the shadow price of product quality additions is almost the same as that for process efficiency; this can be observed in Appendix 4.1.

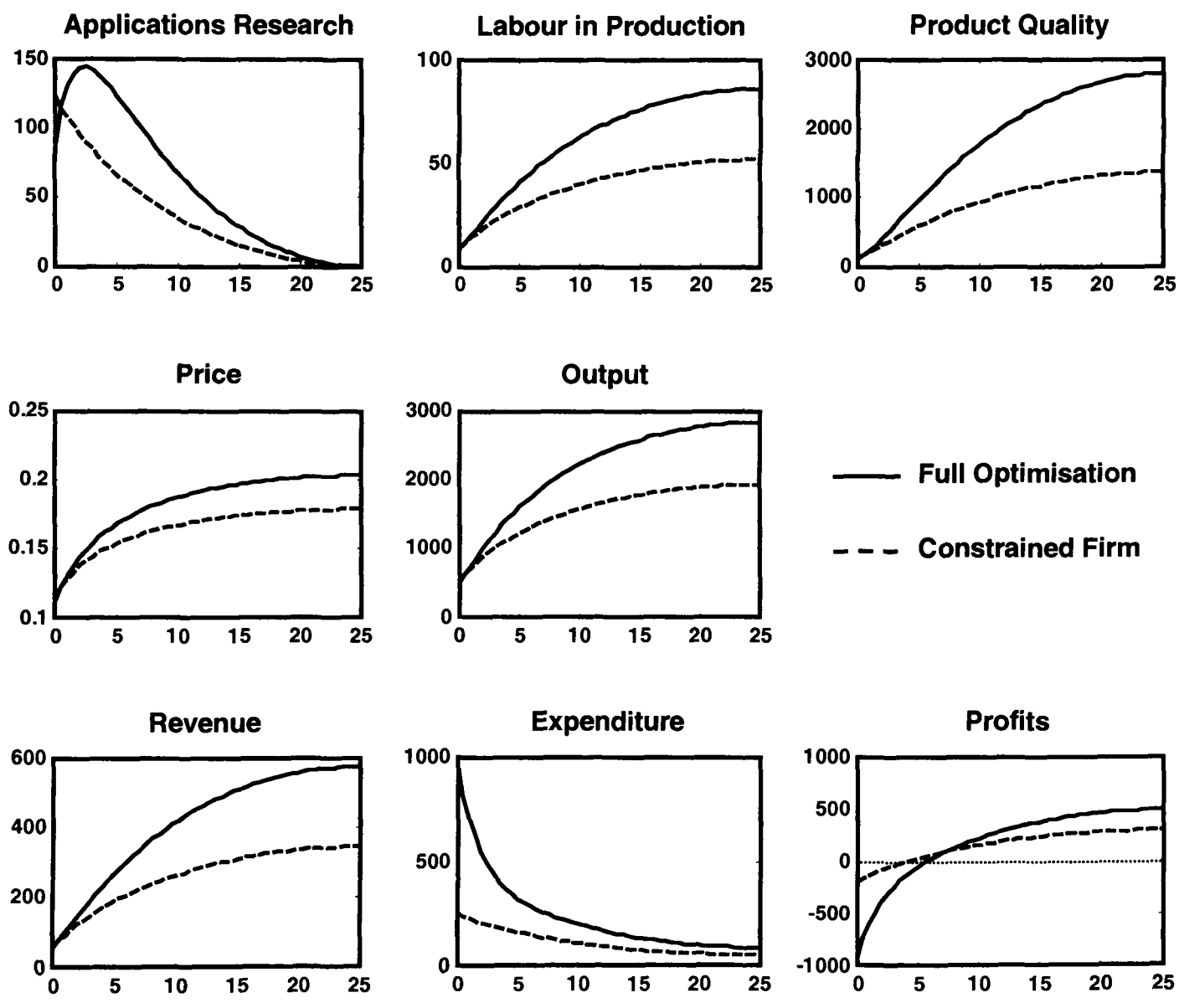
Figure 4.3 indicates that the price is approximately the same at the beginning of the planning span. It rises gradually as the firm offers products of improved quality, to be about double the original level by the terminal time. When the firm is improving its production technology, a sharp decline in the price is observed close to the initial time, before the bulk of R&D has been expended, and then the price stabilises at around a tenth of the original price. Employment in manufacturing is greater with process innovation throughout the development cycle. The improving labour productivity when the firm is orienting its R&D towards bettering its manufacturing processes implies that output in this case becomes several orders of magnitude greater by the terminal time. Significantly, a comparison of the variables having to do with the financial condition of the firm reveals a close proximity between current and cumulative profits, even though the wage bill and revenue are considerably higher in the process specification. This lends some support to the hypothesis that the outstanding difference between these alternatives of technical advance concerns the instruments to do with pricing and the scale of production, which may affect the firm's balance sheet, although not necessarily its profitability.

Another revealing experiment is to contrast the decisions displayed in Figures 4.1 and 4.2 to those of a firm that is constrained to employing applications-oriented researchers – in other words, one lacking exploratory research as an instrument of innovation. To generate these results, one must do away with the differential equations describing the evolution of ψ_S and S and set the state of knowledge at the same level $S(t_0)=S(t)=100$ for the whole span of planning, leaving the other parameters at their benchmark values.[7] This constrained problem (the results of which are plotted in Figures 4.4 and Figure A4.1 in Appendix 4.3)

[7] The veracity of these assertions can be shown formally by deriving the system of differential equation corresponding to the necessary conditions of an optimisation problem in which the firm only has two control variables, the number of workers employed in manufacturing and applications-oriented research workers. This exercise also shows that the benchmark parameter values satisfy the sufficient conditions for a global maximum.

generates a solution exhibiting a strictly decreasing expenditure on applications R&D; i.e., one in which maximum spending occurs at the start and then declines monotonically. This result is readily explained: the firm has no reason to 'delay' its expenditures on applications R&D. It is interesting that the firm with no exploratory R&D would hire *more* researchers to develop applications close to the initial time t_0. All the same, the firm investing in exploratory R&D, which sees the marginal productivity of its applications-oriented R&D increase, would ramp up its applied research expenditure to levels above that of a constrained firm after a short period of time.

Figure 4.4 Product innovation with and without exploratory R&D



With regard to the indicators of financial performance, the figures reveal that the wage bill is sharply cut at the outset in the constrained specification and remains lower later on,

regardless of whether the firm is improving the quality of the product it offers or the efficiency of the production processes it employs. This pattern conceals two separate effects: first, that the firm is not funding any exploratory research and less applied research; and second, that the workforce employed in manufacturing has been scaled down. The firm that is prevented from allocating research workers to exploratory research activities forgoes some revenue as well, because it is selling fewer units of the commodity at a lower price. The end result is a decline in cumulative profits; unless this was the case, it would have been optimal for the firm funding both research activities to suppress its exploratory research altogether in order to gain additional profits. But it is worth noticing that this reduction hides a more complicated pattern in the evolution of current profits: early on profits are higher (losses are lower) for the firm that only can fund applications-oriented R&D but, later, they are below what is earned by a firm simultaneously undertaking exploratory research. As a result of this 'swing' in current profits, the break-even point is closer to the initial time.

The different intensity of these research programmes has as a consequence that the path for the state of product quality or process efficiency is uniformly lower and strictly concave (rather than S-shaped) when the firm is restricted to funding applications-oriented research.[8] The same observations are true with respect to the labour force engaged in production and of output, which again are lower and strictly concave in the constrained simulation. The preceding results help to explain why the rate of change of prices is not as great. In comparison to the price of the product sold by a firm that may increase its fundamental knowledge resources, the price of a firm confronted with a constant state of knowledge will be: lower throughout the planning horizon when it is improving its product

[8] Equally, the shadow prices of product quality or process improvements are uniformly lower when no exploratory research is envisioned. This is to be expected since the marginal productivity of applied researchers is tied to the state of knowledge, which in this case is fixed, not increasing.

design, reflecting consumer's lessened willingness to pay for a consistently inferior product; marginally higher when it is improving its production system.

4.5 Comparative dynamic analysis

The comparative dynamics results in this section were derived by adjusting each of the model's parameters to values above and below the benchmark, leaving the other parameters as in the benchmark case, and running new simulations to obtain the optimised trajectories for the state, co-state and control variables. The changes to the solution that result from these variations are synthesised in Tables 4.3 and 4.4. The discussion that follows attempts to construe plausible interpretations for the re-optimised research and production programmes, by looking in turn at the parameters related to product market conditions, technical and research capabilities, labour market conditions and the discount rate.

Table 4.3 Comparative dynamics for product innovation

Higher:	ψ_Z	ψ_S	Z	S	N_A	N_E	L	p	q	Rev[1]	Exp[1]	Π[1]
χ	+	+	+	+	+	+	+	+	+	+	+	+
ε	-	-	-	-	-	-	+	-	+	+/-	-/+	-
α	+	+	+	+	+	+	+	-	+	+	+	+
ϕ	-	-	+	+	-/+	+	+	+	+	+	+	+
w	-	-	-	-	-	-	-	+	-	-	-	-
w_A	+	-	-	-	-	-	-	-	-	-	-	-
w_E	+	+	-	-	+/-[2]	-	-	-	-	-	-	-
w_A & w_E[3]	+	+	-	-	-	-	-	-	-	-	-	-
δ	-/+	-/+	-	-	-/+	-/+	-	-	-	-	-	-
$Z(0)$	-	-	+	-	-	-	+	+	+	+	-/+	+
$T(0)$	+	+	+	+	+	+	+	-	+	+	+	+
$S(0)$	-	-	+	+	+	-	+	+	+	+	-/+	+
t_1	+	+	+	+	+	+	+	+	+	+	+	+

[1] 'Rev' stands for Revenue, 'Exp' for Expenditure, 'Π' for *Cumulative* Profits.

[2] '+/-' denotes an initial increase and a subsequent decrease in the variable's evolution; likewise, '-/+' denotes an initial decrease and a subsequent increase in the variable's evolution.

[3] 'w_A & w_E' refers to an across the board rise in wages for research workers.

Table 4.4 Comparative dynamics for process innovation

Higher:	ψ_T	ψ_S	T	S	N_A	N_E	L	p	q	Rev[1]	Exp[1]	Π[1]
χ	+	+	+	+	+	+	+	-	+	+	+	+
ε	-	-	-	-	-	-	+	-	+	+/-	-/+	-
α	+	+	+	+	+	+	+	-	+	+	+	+
ϕ	-	-	+	+	-/+	+	+	-	+	+	+	+
w	-	-	-	-	-	-	-	+	-	-	-	-
w_A	+	-	-	-	-	-	-	+	-	-	-	-
w_E	+	+	-	-	+/-[2]	-	-	+	-	-	-	-
w_A & w_E[3]	+	+	-	-	-	-	-	+	-	-	-	-
δ	-/+	-/+	-	-	-/+	-/+	-	+	-	-	-	-
$Z(0)$	+	+	+	+	+	+	+	-	+	+	+	+
$T(0)$	-	-	+	-	-	-	+	-	+	+	-/+	+
$S(0)$	-	-	+	+	+	-	+	-	+	+	-/+	+
t_1	+	+	+	+	+	+	+	-	+	+	+	+

[1] 'Rev' stands for Revenue, 'Exp' for Expenditure, 'Π' for *Cumulative* Profits.

[2] '+/-' denotes an initial increase and a subsequent decrease in the variable's evolution; likewise, '-/+' denotes an initial decrease and a subsequent increase in the variable's evolution.

[3] 'w_A & w_E' refers to an across the board rise in wages for research workers.

Before discussing the results, it is advisable at this point to emphasise that the comparative dynamics need to be conceived as a mapping from different parameter values to optimal trajectories, broadly fixing limits in a "solution surface" spanning the full (admissible) parameter range. Since one is dealing with 'snapshots' of the solution surface corresponding to a specified set of conditions faced by the firm, the arguments regarding the changes in the optimal evolution to parameter variations cannot be made in terms of cause and effect. This is the explanation to some (apparently) puzzling findings; for example, why modifications in ϕ or $S(0)$ "result" in the shadow price for applications and knowledge moving in the opposite direction to the employment of applications and exploratory research workers.

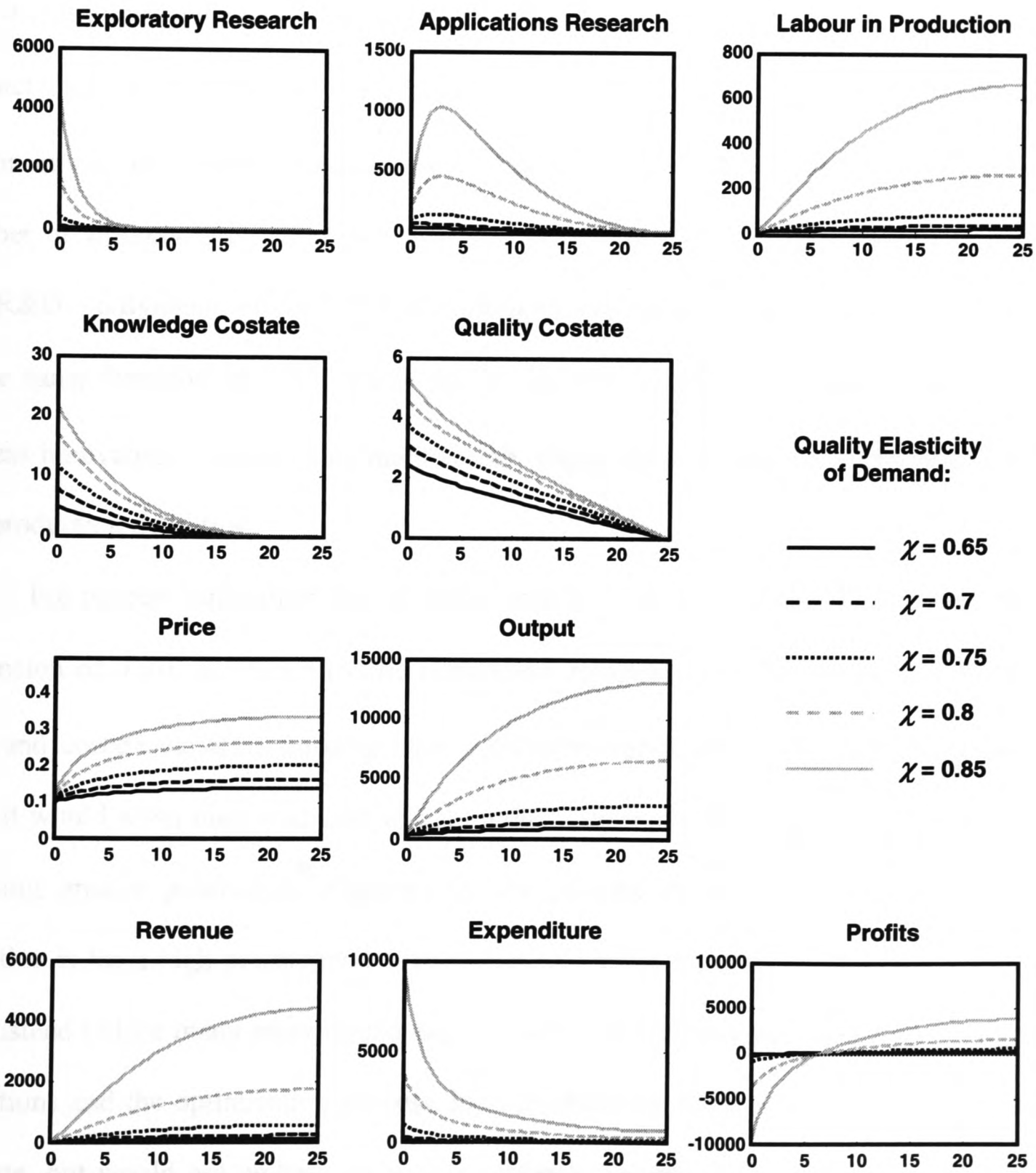
4.5.1 Product market conditions

In the model, the responsiveness of the final product market to quality and price is governed by the elasticities of demand χ and ε respectively. Comparative dynamics with respect to χ suggest that the research efforts of a firm that is improving its product design would be more intense in markets where consumers are more responsive to quality. Accordingly, the evolution of product quality is steeper the greater is the quality elasticity of demand, and the firm can exploit this by increasing its output without depressing consumer' willingness to pay as much. The end result shows a steeper evolution in price also. Expenditure, revenue and cumulative profits are higher the more elastic demand is to quality. See Figure 4.5, which depicts the optimal trajectories of the each variable when a product innovating firm faces different values of χ. In respect to process innovation, since product quality is fixed throughout, a positive change in χ simply scales up the demand function. A profit maximising firm can take advantage of a 'larger' market by upgrading its production technology more rapidly, and pushing the product onto the market with lower prices, again making a net cumulative profit.

When demand is more price elastic (higher ε), a firm conducting product innovation is more limited in its capacity to extract resources from consumers through higher prices. The price path is uniformly lower, and a greater share of revenue increases is earned from expansions in the scale of production as more workers are employed in manufacturing. The adjustments in these optimal time paths take the same direction when the firm is improving its manufacturing process. It has long been recognised that some product market features may be more or less conducive to innovation. The comparative dynamics results indicate that a profit-maximising firm will find it in its interest to spend less on both classes of R&D in price elastic markets, as one would expect a priori. Despite the observed decline in research

expenditures, their 'timing' remains roughly the same, in the sense that exploratory research is exhausted in the same 'period' under different values of ε and, consequently, the peak in applications-oriented research occurs at the same point in time with each simulation. The optimal controls determined in price elastic environments result in reduced spending on wages and revenue overall, and a negative effect in terms of cumulative profits, in both the product and process specifications of the model.

Figure 4.5 Quality elasticity of demand and product innovation



4.5.2 Production and research capabilities

The production possibilities of the firm are determined by the shape of its production function, summarised by the parameter α, and the initial state of its production technology $T(0)$. The variations in the optimised research and production plan when α is higher are as expected: when the firm is nearer to having constant returns to scale, its benefits from undertaking product or process innovation are greater. This is reflected in higher R&D expenditures, output, and profitability. When $T(0)$ is higher the ensuing profile depends on the kind of innovation under consideration. Since the efficiency of manufacturing is fixed for product innovation, a higher $T(0)=T(t)$ induces a one-off hike in the firm's equilibrium labour inputs and output. Because the market that can be served by employing a given number of workers is enlarged, the firm also finds it optimal to direct additional resources into R&D – a dynamic effect. Revenue, aggregate expenditures and cumulative profits move in the same direction as $T(0)$. These results are very similar to a higher χ in the case of process innovation, outlined a moment earlier, triggering a scaling up of the firm's research and production activities.

For process innovation, the "solution surface" is more uneven when moving along the dimension of $T(0)$, as shown by the remarkable re-optimisation in the path of the control, state and costate variables ensuing from a different initial production technology (in Table 4.4). It would seem that, when the initial productivity is already high, the marginal value of attaining greater production efficiency is low relative to its cost. The response to this valuation is for a high productivity firm to favour a lower schedule of research expenditures and instead to hire many more employees to work in manufacturing. As a result of the initial conditions and the optimisation process, high productivity firms would have more sales and revenue, but would not embark on such an intense research process. This suggests that the

rather banal observation that "more profits accrue to a more efficient firm" may have an interesting dynamic element linked to the firm's equilibrium research intensity. Part of the gain in a product monopolist's cumulative profits comes from being better endowed technologically to start off with. But in the context of a market demand that only is willing to absorb so much production at a given price and quality, a large proportion of the additional profits are earned by channelling spending towards production instead of paying for costly additions to its knowledge base or to improve its production system design.

Mutatis mutandis, this argument is equally valid for a product innovating firm that starts out with an improved product design, as can be observed from the comparative dynamics with respect to $Z(0)$. Again, the firm appears to have weaker incentives to fund research and instead steps up its employment in manufacturing at the going market wages. Cumulative profits are higher, as a consequence of the initial state of its product quality and of the firm's reduced R&D spending in equilibrium. Finally, it can be seen from Table 4.4 that, in the case of a process innovating firm, a superior product design acts as a scaling up of demand.

The optimal control trajectories decided upon by the firm when the initial boundary values for the state of product quality or process efficiency are varied may provide a complementary explanation for some empirical observations related to the development of innovative industries. Consider, as an example, Esposito and Mastroianni's (2002) empirical study of the changes that have lately taken place in the personal computer (PC) industry. They find that as PC's became more advanced in their capacity to process information, as measured by strictly technological indicators such as RAM size and processor speed, the price-performance relationship broke down: while the price of PC's was closely correlated to these performance (quality) indicators when they were commercialised in the late 1980's, this was not the case later on. They suggest that this finding can be explained by a policy of price

discrimination by firms between two market segments that entered the industry at different times. The original "innovator user (computer science professionals, office automation users, connoisseurs) interested in the performance, and the conservative user (mass user) interested in the cost,"[9] the latter segment being more price elastic – or in any case composed of consumers that were attracted by novel features of PC's, such as multimedia applications.

A different justification for this statistical finding can be made by extrapolating from the comparative dynamics. Suppose a firm early on in the PC industry had as yet only low performance goods (low $Z(0)$), for which there are high incentives towards product innovation and low incentives for process innovations – since the latter leads to profit gains in proportion to the market size already served. Such a firm would presumably prefer to improve its product design, and the analytical and numerical results suggest that it is then optimal to increase the price at a proportional rate. There may come a point, however, when the product quality Z may be large enough to justify a shift in the direction of process innovation. When this occurs, the firm would change its pricing scheme, and a sharp decrease in the price is to be expected. The pattern of innovation and pricing that has been suggested not only explains why the relationship between price and quality could break down, but also the nonmonotonic price evolution observed by Esposito and Mastroianni, first increasing at a slow pace and then sharply decreasing later in PC's "technological life cycle." Of course, to corroborate this claim one would need to extend the scope of the model, by explicitly allowing the firm to design a research plan in which it sequentially supports the two modes of innovation. In the last section we give an indication of the potential for this line of research.

Turning to the firm's research capabilities, the relevant parameters are the elasticities in the production functions for knowledge (ϕ) and applications (υ_Z and υ_T), and the initial value for the state of knowledge $S(0)$. As remarked in Section 4.3.3, which evaluated the

[9] Esposito and Mastroianni (2002), p. 44.

sensitivity of the results, a comparative dynamic analysis with respect to these parameters is hampered by the nonlinearity of the differential equations. In the case of ϕ, solutions involving complex numbers can be avoided by considering variations that are close to the benchmark value. One can then conclude that a firm with more productive exploratory research activities (ϕ closer to 1) has, as one might expect, an incentive to allocate further resources towards exploratory R&D. A steeper evolution in the state of fundamental knowledge has a derivative effect on applications R&D, which is generally higher, as a result of the greater marginal productivity of this activity.

Whilst the overall intensity of commercially-oriented R&D on applications may be greater when the firm's exploratory research is more efficient, this does not mean it will be uniformly higher over the planning horizon: the results suggest that the firm reschedules the bulk of its applied research until after the exploratory researchers have made an important headway, delaying the former until the fundamental knowledge resources are available. Indeed, in the comparative dynamics for product and process innovation, the firm hires fewer research workers to search for applications at the outset. It is only after the exploratory research is under way that the firm accelerates its search for applications, which will attain a more elevated 'peak' the greater is ϕ. This progression is accompanied by higher output, revenue and cumulative profits. The puzzling result that R&D expenditures and their shadow prices move in opposite directions may be explained, as mentioned above, by the fact that this only is a slice in the 'solution surface.'

Changes in υ_Z and υ_T, the share of applications-oriented research labour in advancing the product design and the production technology respectively, generate a convoluted re-optimisation of the optimal paths, even for those values of the parameter for which the results are real numbers. Table 4.5 illustrates the point: as knowledge becomes more important in improving product quality (υ_Z goes from 0.5 to 0.75), there is a complete

rebalancing of the research efforts, production, etcetera. Applied research 'swings' around, being much larger at the start but lower later on, and the path for exploratory research is uniformly lower. It is hard to make a definitive assessment about the precise influence of this parameter on the equilibrium determination of the firm's controls, the evolution of fundamental knowledge resources at its disposal or the practical advances it makes in terms of product quality or the production technology. As we argued before, the mathematical rationale for this negative result is that the solution is particularly sensitive in the dimension of υ_Z and υ_T because the parameters are pivotal to the nonlinearity exhibited by the system of ordinary differential equations resulting from the necessary conditions for a maximum.

Table 4.5 Additional comparative dynamics for product innovation

Higher:	ψ_Z	ψ_S	T	S	N_A	N_E	L	p	q	Rev[1]	Exp[1]	Π^1
υ_Z	-/+	+/-	+/-	+/-	+/-	-	+/-	+/-	+/-	+/-	-	-

Lastly, the results from the simulations investigating variations in the initial state of fundamental knowledge $S(0)$ can be explained in similar terms to the simulations studying variations in initial quality in the case of product innovation, or in production efficiency for a firm undertaking process innovation. That is, they suggest that a firm endowed with a superior store of knowledge or better-trained research workers has less of an incentive to push its fundamental knowledge frontier further forward. It will select to spend a higher amount on applications-oriented R&D, given this is more productive in absolute terms and more profitable compared to exploratory R&D. There is therefore a substitution effect working between different classes of knowledge, as their availability changes without their development costs doing so. This effect brings about a quite different research programme depending on the value of $S(0)$. As one would expect, a firm that has greater research capabilities from the outset has a steeper rate of incline in the state of product quality or

production efficiency, depending on the objective of its research; as a result, its revenue stream is enlarged, as are its profits.

4.5.3 Labour market conditions

The highly stylised labour market specified in this and the preceding chapter assumes that employees may be hired at the equilibrium market wage w to work in manufacturing, whereas researchers need to be paid w_A when they are allocated to commercially oriented R&D and w_E if they are assigned to exploratory research. Up to now, all the comparative dynamics results have been derived under the restriction that $w_A = w_E$. We will retain this assumption for a moment in order to examine how the firm's optimisation responds to an environment where the relative wage of research workers vis-à-vis the workforce employed in manufacturing is higher or lower.

If w_A and w_E are higher, the results from the comparative dynamics indicate that the firm employs fewer researchers in equilibrium, as one might well expect. As a result of these control trajectories the quality offered, or the state of its production technology, depending on whether we consider product or process innovation, are lower at each point in time. A slower growth rate of product quality limits the output the firm can sell at a given price, and in consequence it retains a smaller workforce in production as well. The labour force is smaller for a firm improving its production processes as well, in this case as a direct result of the lower marginal revenue product of labour. In either specification, these effects push down the time paths for revenue and expenditure. The combined effect on cumulative profits shows this to be lower the higher researchers' wage. The pattern that surfaces from the comparative dynamics with respect to the wage of workers in manufacturing is that a higher wage is akin to more pronounced decreasing returns to scale in the production function for goods. In both

cases, the potential gain from a given investment in R&D is lessened, and this is smaller in equilibrium, as are expenditure, revenue and cumulative profits.

Suppose, instead, that the going market wage for research workers is different depending on the activities they will undertake in the firm. The wage differentials may have their origins in various features of the labour market. For example, they could arise if there was a greater relative scarcity of the specific skills needed to properly carry out one of the two research activities.[10] The general characteristics of the numerical solution are the same in this setting; i.e., the evolution of applications research will be single peaked, the path of the product's quality (or the state of the production technology) can be S-shaped or concave, depending on the amount of exploratory research undertaken, etcetera. And, at first blush, it appears that higher labour costs for either exploratory- or applications-oriented research lead to an adjustment of the optimal control, state and costate variables that is almost exactly the same to that outlined above, in which we considered an equal increase in w_A and w_E. But there are distinctive elements to the comparative dynamics when only one of the two types of research becomes more costly.

The first of these can be directly observed from Tables 4.3 and 4.4: higher values for w_A depress applied R&D expenditures by the firm across the entire planning horizon, whereas a higher w_E may elicit a small hike in the resources allocated to this mode of research close to t_0, followed by a more substantial dip later on. This phenomenon can be explained in the same terms as our earlier discussion of variations in the efficiency of

[10] And, whilst the premium on these skills may originate on the supply-side, it can be exacerbated when there is a bias in the demand for research workers from other firms or public institutions; for instance, if they are prone to funding a certain form of R&D more than others. Indeed, if these conditions are very acute, the research skills demanded by an innovating firm may be in such short supply that it must explicitly consider the wage-effects of implementing its research programme – in other words, it will face an inelastic labour supply curve. Such an eventuality is not considered in this analysis.

exploratory researchers (ϕ), by recognising that a profit-maximising firm delays or brings forward the bulk of its applications oriented research to get the greatest impulse from its expanding knowledge base. When the relative cost of its search for fundamental knowledge is higher, it will cut back its funding of exploratory research. Given the fixed horizon within which it can collect the gains from the improved product or production technology designs, it is then optimal for the firm to front-load its applied R&D spending.[11]

This evaluation of the relationship between the constellation of labour costs to the firm's optimal research plan may be brought to an end with a more general comment bearing on the complementarity of labour inputs in this model. The comparative dynamic analysis indicate that upward changes in either w, w_A or w_E generate solutions in which the research intensity and production is generally reduced. There is some inter-temporal substitution of research activities, but if one takes the aggregate wage bill of research workers at any point in time, it turns out to be consistently lower, as does the cumulative spending on each kind of R&D. This suggests that the substitution effect that may arise in static settings, whereby a higher relative wage for one group of workers can lead to more employment of the other groups, is far outweighed by dynamic considerations pertaining to the links between: fundamental knowledge resources and the marginal productivity of commercially-oriented research on applications; and research efforts taken as a whole and the firm's potential market (in the case of product innovation) or its ability to supply a standard good at a given cost (in the case of process innovation).

[11] The other points of contrast follow from a closer inspection of the simulations. This inspection reveals that the direction of the adjustment may be the same in the other variables, but there are significant differences in the relative size of the changes to equal variations in these parameters. Compared to the same variation in w_A, when this is in w_E the solution is prone to a more marked shift in exploratory research and less so in respect of applied research; connectedly, it exhibits greater variability in the state and costate variables associated to fundamental knowledge additions, but a narrower range of variation for the state and costate variables linked to the product quality or the state of the production technology.

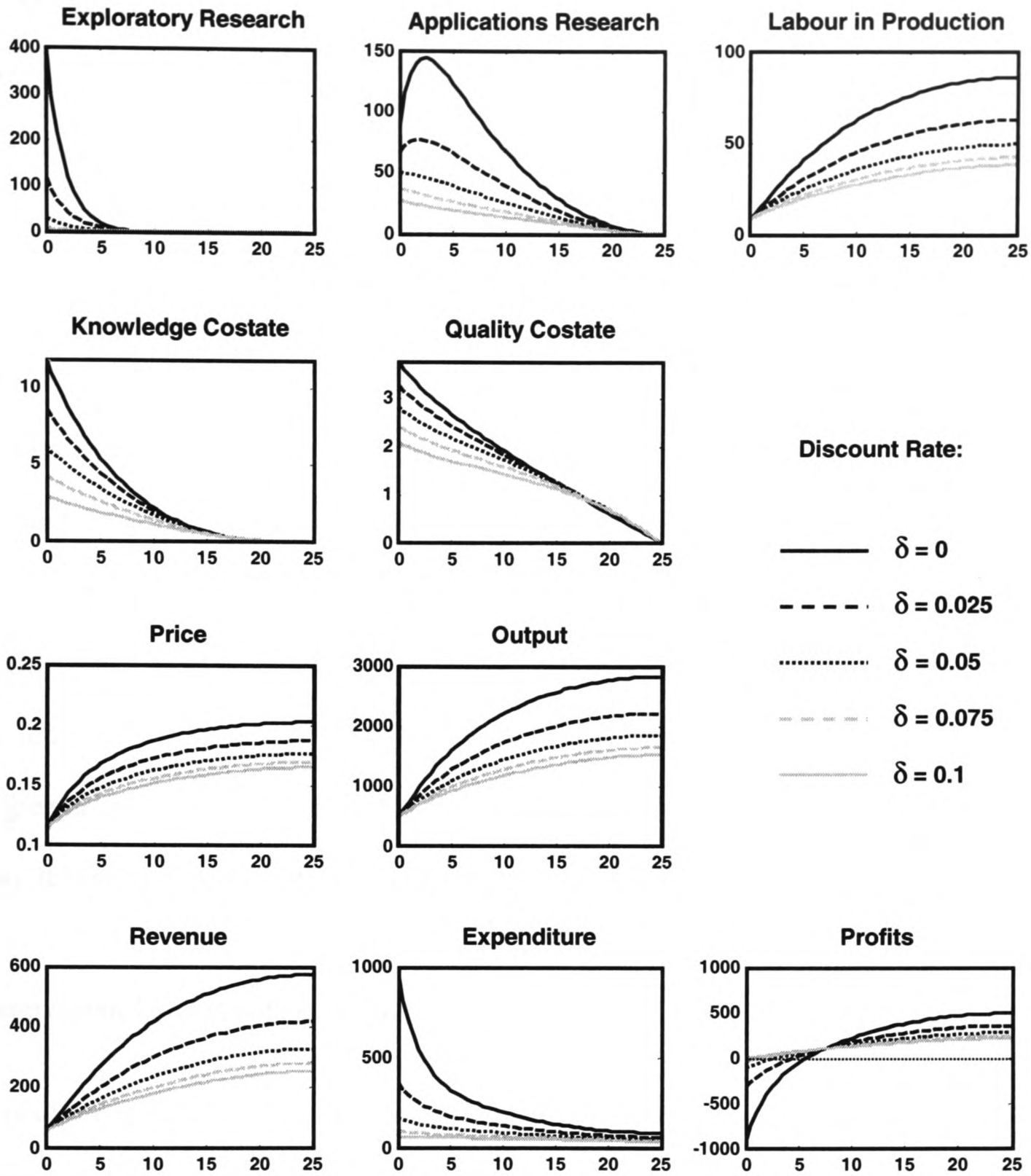
4.5.4 Discount rate

Up to this point, all of the results presented have been derived under the assumption that the firm does not discount its stream of profits. But there is reason to think that discounting will be central to the assessment of R&D investment projects, as these are typically undertaken over long periods of time, and this section aims to clarify the manner in which a positive discount rate influences the scale of the firm's R&D spending, and the share of the resources supporting each mode of research. It seems convenient to pre-empt the discussion of these results by noting that discounting does not affect any of the conclusions so far reached, and that the general features of the solution resurface in the simulations with positive discount rates. This is significant, because it lends an extra degree of robustness to the description of the optimal dynamic profiles for research and production and its underlying foundations. The commonalities also will allow the remainder of this section to focus specifically on the comparative dynamics aspects of discounting and thus to avoid making a repetitive account of the general trends in the control, state, and costate variables.

Figure 4.6 depicts the optimal trajectories that result with values of δ ranging from the benchmark of zero to δ=0.1 for product innovation; Tables 4.3 and 4.4 present a summary of the results for both product and process innovation. This range is sufficiently ample to get a good overview of the range of outcomes that are supported as a maximum to the problem.[12] The simulations show that higher discount rates depress spending on R&D, and accordingly lead to a solution in which the firm cuts back its production, with which it earns far less revenue also, even though the implied price of its product is higher.

[12] To recognize what different values of the discount rate stand for, notice that, as the simulations are run over an interval of 25 'periods,' the present value of 100 units of revenue earned at the terminal time t_1 equals roughly: 50 units at the initial time t_0 when δ=.025; 30 units when δ=.05; 15 units when δ=.075; and 10 units when δ=.1.

Figure 4.6 Results of the simulations for product innovation with discounting



A noteworthy aspect of the time paths of exploratory and applied research is that higher discount rates seem to press down more forcefully on the former. That is to say, the trajectories followed by the research expenditures do not fall down uniformly, a higher discount rate first 'pushes' the firm into selecting the "boundary" in which it only funds applications research and then to the "boundary" with no research at all (this kind of expenditure is negligible for values of δ in excess of 0.2). As a result, the observations made

in other sections about the changing time profile of applied research are readily observable in Figure 4.6: the peak flattens and is closer to the initial time if the discount rate is higher. For values of the discount rate above 0.05, almost no exploratory research is undertaken, the peak in applications-R&D is at t_0 and it is monotonically declining throughout the planning interval.

Finally, note the one difference that exists between the model with and without discounting. This is that the evolution of the costate variables (ψ_S, ψ_Z, and ψ_T) is not necessarily convex to the origin, as we proved them to be in the formal derivation of the undiscounted model in Chapter 3. Although they still are decreasing over time, they may not do so at a decreasing rate. And, a closer examination of their time paths for different values of δ, shows that they 'twist around' and cross each other close to the terminal time. In other words, the marginal valuation of additional knowledge or applications in profit terms seems to be greater for a firm with a higher discount rate towards the latter part of its planning horizon. It is not clear what (if any) economic implications obtain from this particular result.

4.6 Innovation in a volatile environment

As a preamble to last section's analysis, we mentioned some limits to the interpretation of the comparative dynamics exercise, in that the optimal time paths obtained for different parameter values represented 'slices' in the solution surface. Moreover, it would be incorrect to extrapolate from the comparative dynamics results, and suggest that a given change in the prevalent conditions *midway* through the firm's optimal programme would generate this or that kind of adjustment in the optimal trajectories in the time that remained. Comparative dynamics hint at how an innovating firm confronted with an externally-driven and unpredicted shock might respond, but to be positive one must explicitly take account of its internal production and research capabilities at the time the shock occurs, as these will have

benefited from its past investments in knowledge-providing research activities. Crucially, the irreversibility of the firm's R&D investments imply that its new research and production plan must be premised on the states of fundamental knowledge, product quality design and production technology when the exogenous shock takes place. The scale and direction of its R&D, which was optimal with the initial set of conditions, may now be a burden on the firm's choice and seem much less valuable than originally.

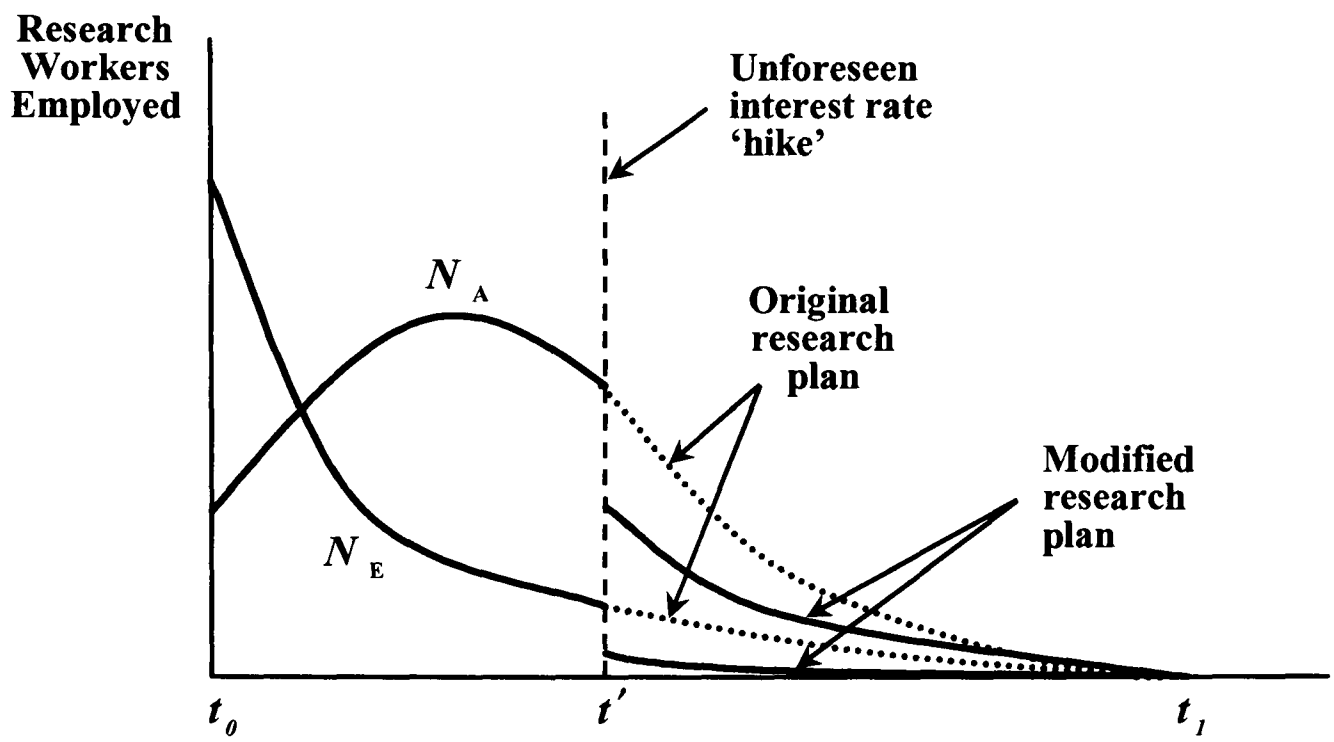
To make matters more concrete, consider a firm that has disbursed the greater part of its planned spending on exploratory research, and an intermediate amount of its investment on applications-oriented research, and is suddenly faced with an unexpected hike in the interest rate that causes it to raise its subjective discount rate.[13] Based on the comparative dynamic analysis, a sensible forecast is that the firm would cut back the scale of its future research activities, and that a relatively stronger reduction is likely to be made in respect of exploratory research expenditures. The expected profile of the trajectories followed by R&D spending before and after the interest rate shock is depicted in Figure 4.7.

Whilst that prediction turns out to be true in the analysis shown below, a priori one would be hard-pressed to be more specific and hazard a guess about, e.g., the extent of the fall in R&D spending compared to the optimal path that would have been selected if the high interest rate had been prevalent through the entire planning horizon, or the manner in which the firm could restructure its production programme to minimise the effects of the shock on its subsequent profits. And, in any case, it is doubtful that there will be a 'hard and fast' rule to the firm's response. The goal of the firm's R&D (to improve its product quality or processes) and the nature of the shock are likely to have a bearing on it's correction of the

[13] In view that the interest rate is a central component of discounting estimates, the terms discount and interest rate will be employed interchangeably in the discussion, although there may be a host of other factors impinging on the discount rate having to do with, e.g., shareholders impatience for returns, or imperfections in the financial market.

optimal controls – whether this is mostly or just a level-effect in a given variable, or whether it requires significant changes to the optimised growth rates.

Figure 4.7 Re-optimising research activities in the face of an interest rate change



Shocks may differ in their source, timing, direction and magnitude. Apart from the discount rate, there may be shocks that originate in the product market, the labour market, or in the scientific knowledge domain relevant to the firm's applied research activities. The last would reflect the discoveries made in universities and other institutions that find their way onto the public domain in scientific data and information. These research processes are exogenous to any action of the firm in question, their motion is governed by different considerations to those of a profit-maximising firm, and these considerations may differ in their importance amongst, e.g., fields of knowledge, countries, and the historical period in question.[14]

Another crucial aspect is the moment in time when the shock takes place; or, to be more exact, the time in relation to the product or process lifecycle that has been identified,

[14] To give a topical example, Paul A. David has argued that the intellectual property right system emerging in developed countries (designed to encourage industrial funding of R&D activities) will have undesirable consequences for exploratory research conducted in 'open science' institutions; see David (1993a, 2002, 2003a, and assorted others listed in the bibliography of those papers).

and the research programme that provides the thrust for this evolution. Is the shock close to the initial time, which would allow the firm greater flexibility to re-plan the scale and mix of its research? Or is it later on, when the bulk of its exploratory- or even its applications-oriented research has been completed? Lastly, and all-important, the direction of the shock, because upward and downward variations in a parameter may elicit asymmetric responses in the context of irreversibility of R&D investments, and its magnitude. Plainly, the types of volatility that can be conceived are endless,[15] and it is not the purpose of this study to investigate them exhaustively. Instead, the rest of this section introduces a numerical "continuation method" that can be employed to investigate this class of variations, and illustrates the method's power by applying it in three economically significant scenarios: the hike in the interest rate exemplified above; an increase in the market wage of researchers; and an external improvement in the state of knowledge.

The continuation method consists of working out the solution to two optimal control problems: the original problem facing the firm, and another for the subinterval going from the time of the shock, which is denoted t', to the terminal time t_1, after making the appropriate changes to the initial conditions and/or parameters. If one is examining a change to a parameter such as δ or w_A, the initial conditions corresponding to the continuation problem are the optimised values of the state variables computed for the original problem at t'; that is, $S^*(t')$ and, depending on the mode of innovation, either $Z^*(t')$ or $T^*(t')$. The simulation is then run for the interval $[t', t_1]$ with the new parameters, but the original transversality

[15] Not only do they include all the combinations in the basic 'dimensions' of a shock, but also more intricate scenarios in which the relevant conditions change in several dimensions at once, or in which the firm is bombarded with a series of different shocks, which may or may not be in the same direction. And there would be nothing fanciful about wishing to model recurring multi-dimensional shocks; indeed, there is reason to think that this is the kind of volatility that occupies centre stage in the complex and evolving panorama that firms must scan as they make their investment and production decisions. But in order to examine that class of shocks it would be preferable to work with a calibrated model so as to ascertain whether the correction of, say, exploratory R&D spending to a simultaneous 2% rise in interest rates and a 10% fall in market wages, is positive or negative overall.

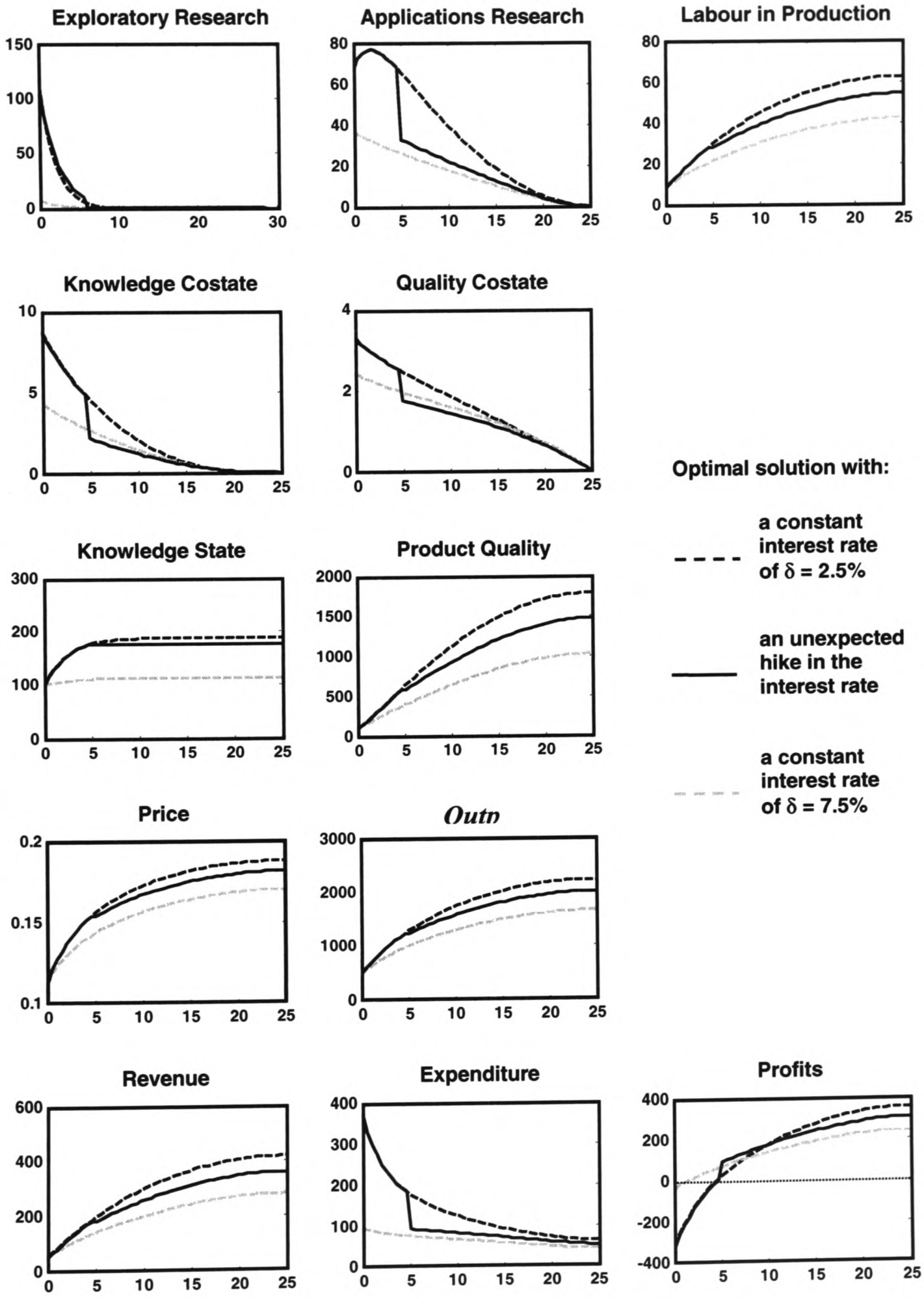
conditions. To model a discrete change in the state of knowledge of 'size' Δ, the initial value of $S(t)$ in the continuation problem would be $S(t')=S^*(t')+\Delta$. The other state variables would have a starting value of either $Z^*(t')$ or $T^*(t')$, whereas the parameters and transversality conditions would be left as in the original problem.

Clearly, these discrete *and* unanticipated variations in the parameters or the state variables do not jeopardise the existence of a solution for the continuation problem, which is simply a duplicate of the problem for the entire planning horizon, and satisfies all of the sufficient conditions that were determined in Chapter 3, and outlined in Section 4.3. This is not necessarily true of a firm that knows about the future shifts in the parameters or state variables *ab initio* because the jumps introduce convexities to the revenue function or to the evolution of the state variables that, if strong enough, could cause the solution to 'explode'. At any rate, the firm would not generally select the same controls when there are jumps in the parameters or state variables. To explicitly analyse that class of dynamic problems requires a nontrivial extension to the maximum principle; Leonard and Van Long (1992, pp. 310-32) outline the necessary and sufficient conditions for a problem with jumps in the state variables, conditions originally supplied by Vind (1967).

Figure 4.8 charts the computed impact of an upward shock in the interest rate by juxtaposing the solutions to three optimisation problems, where a product-innovating firm confronts: a constant interest rate of 2.5%; a constant interest rate of 7.5%; and a permanent increase in the interest rate from 2.5% to 7.5% at t'=5. Other than the discount rate, the simulations were run with the benchmark parameter values. The re-optimisation that results from this shock has some remarkable features. In regards to R&D spending patterns, investment in exploratory R&D is terminated at t' and there is a sizeable one-time reduction in applications-oriented R&D. The latter then follows a declining time path with a slope in

between what would be optimal with a constant 2.5% interest rate and the flatter progression that would be selected if the rate were 7.5% during the entire planning horizon.

Figure 4.8 Optimal response to an unanticipated hike in interest rates



The firm's optimal reaction needs to be explained in terms of the irreversibility of R&D investments during the 'low interest rate period,' which by t' have generated a greater store of fundamental and practical knowledge than would have existed if the initial interval had been of high interest rates. This feeds into the firm's estimate of the future discounted profits that can be gained by additional R&D spending. In this example, it makes further investments in fundamental knowledge superfluous. But the plentiful knowledge resources already present at t' shore up the marginal productivity of applications-oriented research, and thus the time path followed by this type of research activities is above that corresponding to the constant 7.5% interest rate regime.

The new bearing of the controls directs the evolution of the state variables and balance sheet measures in a rather different direction to what originally had been optimal. Figure 4.8 shows that, once the firm finds itself in the high interest rate regime, it will observe a less inclined path in its employment in manufacturing, output, price and revenue. Since expenditure falls abruptly as the firm curtails its research activities, the profit it registers would improve at the time of the shock; after this time, profits increase at a slower pace and eventually fall behind what the firm would have received if it had faced a low interest rate all the way through. Cumulative profits are evidently greater in the scenario with a stable low rate of interest than when an adverse shock disturbs the research and production programme, but the latter turns out to be greater than the scenario with high interest rates – one imagines that the second comparison could go both ways, as a firm that originally looked forward to a low interest rate invests heavily close to the initial time, and later finds itself with a lower than anticipated return. Several simulations were run to informally check whether similar results emerge when the shock occurs at other moments, either closer to the

initial time or much later on. The results presented in Figure 4.8 and discussed above emerge in these solutions as well.[16]

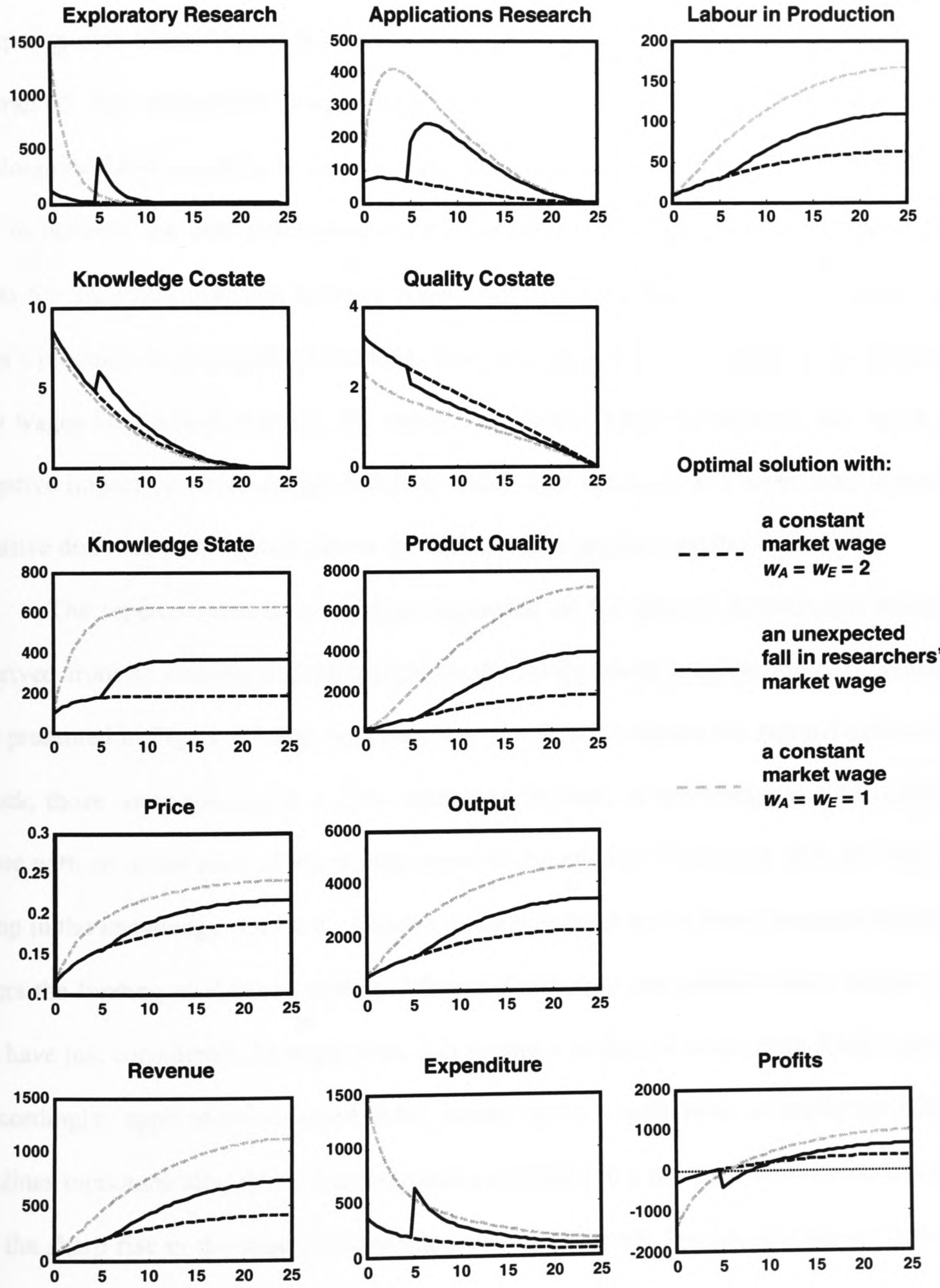
An adverse shock of the type explored in the foregoing example would naturally be expected to inhibit the firm's R&D intensity and result in a squeeze on profits. What is remarkable is the weight of the irreversible investments on the firm's optimisation problem, as it tries to make the most of the design improvements and knowledge resources it has funded up to the time of the shock; the former allows a superior good to be produced without further R&D spending, whereas the latter can be drawn on by its researchers at no further cost for the remainder of the planning horizon. In fact, one can view the readjustment as being mediated by three distinct dynamics: one generated by the higher interest rate, which depresses the present value of further investment; another driven by the relatively better design existing at t', which also has a downward bearing on both classes of R&D; and a third, arising from the improved knowledge base, which has a mixed impact, reducing the benefits of funding additional exploratory research but increasing those from applications research. Only the last of these can be responsible for moderating the fall in applications R&D. As a result, the correction to the firm's controls bears only a limited resemblance to the research and production programme that it would have embarked on in an adverse but stable environment, with high interest rates throughout.

The next two examples inspect the firm's adjustment process in response to two favourable turns of events, in the form of an unforeseen decline in the market wage of research workers and a one-time increase in the state of fundamental knowledge – leaving all other parameters at their benchmark values. In addition to the specific insights that emerge

[16] Interestingly, they suggest that the investments in R&D and the costate variables fall down to meet the same downward trajectories of Figure 1, regardless of when the shock occurs. As a result, the time paths of the state variables after the shock appear to be parallel to each other, with a level difference that depends on the timing of the interest rate hike – the closer it is to the final time, the nearer these are to the curves sketched out with a constant 2.5% rate of interest.

from these simulations, the "logic of sunk-cost hysteresis" can again be seen at play in the firm's correction of its optimal investment behaviour following these positive shocks.

Figure 4.9 Optimal response to an unanticipated fall in research workers' wage



As Figure 4.9 shows, a reduction in research labour costs from $w_A = w_E = 2$ to 1 at t'=5 induces a second 'burst' of exploratory research, starting at t', and a hike in applications research, which peaks shortly afterwards. The upscaling in R&D is accompanied by similar increases in the firm's employment in production and, as a result, output and revenue have steeper growth paths. Notice that exploratory R&D exceeds the level that would have been warranted had researchers wages been low at each point in time; nonetheless, total exploratory R&D spending is lower in the simulation with the shock. The other trajectories fall in between the ones determined by the simulations with low and high research labour costs for the entire planning horizon. Again, this is a reflection of the lower quality of the firm's product and the smaller knowledge base at its disposal at t' relative to the setting with low wages for research workers. By stimulating further R&D investments, this shock has a negative impact on short run profitability, which then increases at a faster rate; overall, the positive direction of the shock allows the firm to make greater cumulative profits.

The repercussions of a one-time expansion of the state of fundamental knowledge (derived from an outside, unpaid source) on the firm's research and production programme are presented in Figure A4.2 in Appendix 4.3. The figure contrasts the optimal paths with no shock, those corresponding to a 25% increase in the state of knowledge at t'=5, and lastly, those with an initial state of knowledge equal to the original benchmark plus the size of the jump in the knowledge state at t'. Clearly, a positive shock in the firm's research capabilities alters the funding of R&D in quite a different direction to the positive labour market shock we have just considered. To begin with, it generates a decline in exploratory R&D, not a rise. Accordingly, applications-oriented R&D jumps up to a new peak immediately and then declines monotonically, rather than increasing gradually to a peak some time after the shock. As the sharp rise in the state of knowledge and the relatively low product design both exert upward pressure on applications research, in this case the latter jumps above the level that

would be funded by a firm that had better knowledge resources all the way through; even so, cumulative spending on applications R&D is not as high when there has been a shock midway through the firm's pre-optimised programme of research. As the fall in exploratory research spending is much smaller than the rise in applications research expenditures, the total wage bill goes up, and profits suffer an initial fall following the shock, and then grow faster than if no shock had occurred.

4.7 Conclusions

This chapter has used numerical methods to analyse a dynamic model of exploratory- and applications-oriented research allocation by a monopoly enterprise. It has found that an innovating firm will set a single-peaked path for its expenditures on applications-research to maximise its cumulative discounted profits and that this peak can be at the initial time or later on depending on the constellation of parameters it is facing. The logic behind this result is that the firm needs to compare the gain from obtaining an improved design early on, which allows it to profit from the inventions for a long period of time, against the value of waiting for a superior knowledge base to be made available by its exploratory research, which increases the marginal productivity of its applied research efforts on a permanent basis. If the firm estimates that 'practical' knowledge embodied in the firm's design of its product or its production system is likely to have a large return, it also will credit fundamental knowledge resources with a high value, and the time profile of its optimal research plan will consist of several distinct phases. In the first of these, a substantial share of the firm's R&D is focused on enlarging its knowledge base. During this time it will steadily increase the intensity of its commercially-oriented research on applications, and this reaches a maximum before the firm stops funding exploratory research altogether. In the remaining time interval, in which it no

longer funds exploratory research, its applied R&D spending falls gradually until this too becomes negligible in size, and the firm ends dedicated solely to production.[17]

By examining the comparative dynamics of the solution, it has been possible to be more precise about the manner in which the product and labour market conditions, firm-level characteristics, and discount rate may impinge on the timing and scale of the optimal research and production programme, and of the relationship between these parameters and the 'boundary conditions,' where only one mode of research is supported, or the firm simply decides not to invest in innovation. The simulations suggest that the magnitude of a firm's investment in R&D would be greater when: demand is inelastic with respect to price and elastic with respect to quality; the production functions for the commodity and the creation of fundamental knowledge are closer to having constant returns to scale; wages, be they of researchers or of employees allocated to production, are lower; the discount rate is low or zero. When the parameters are biased in that direction, the firm also will tend to focus more it its initial R&D budget towards exploratory R&D, and simultaneously to delay the peak of its applied R&D, in which case the evolution of the product quality design or the state of the production technology will follow an S-shaped path.[18] An interesting result revealed by the simulations was that cumulative profits increase when the parameters move in the direction just described, but they do so at the cost of heavier investments and therefore the break-even point for the firm occurs later on. Conversely, if the firm faces a tough product market, is inefficient in its production or research, must pay high wages for its labour inputs, or discounts the future heavily, it would first scrap its expenditure on exploratory research and,

[17] Although cast as a monopoly model, it might be viewed as a growth model in which "layered" research sectors give rise to a "hill-shaped" applications-oriented investment profile. It would be interesting in future work to extend the insights for a larger number of layered sectors. I would like to thank Dr. David Myatt for pointing this out.

[18] This finding is noteworthy because in the previous chapter we conjectured that a profit-maximising firm would probably go through a "technological life cycle" of this kind. We leave it for future work to formally prove the single-peaked property of applications-oriented R&D in the general setting studied there.

if the conditions confronting the firm are more severe, it may forego its research on applications altogether.

Another finding of the comparative dynamic analysis was that a firm that starts out with a superior product design would undertake a more limited research programme to further improve this design, and instead hire more employees to work in manufacturing. The same is true of the research efforts of a firm improving its production technology when this is initially more efficient. We suggested that this could serve to explain why the price-performance relationship has been observed to break down in the PC industry, and to give a reason for the inverted U-shape evolution of price in that industry. To confirm this claim would require an extension to the scope of the model, to explicitly allow the firm to design a research plan in which it sequentially supports product and process innovation. This problem could readily be tackled using the present formulation by thinking about the firm's optimisation problem as being divided into two fixed parts. In each of the two, it must decide whether it directs its commercially-oriented research towards improving its product design or its production technology; e.g., it starts with a research programme to improve its product and then has an option to switch to process innovation, or it may want to continue with product innovation all the way through. To decide between the four alternative research programmes, the firm would need to calculate the present value of the cumulative profits from each and select that which gives the larger sum. A crucial assumption that would need to be addressed in this context is whether the knowledge base can be 'carried over' in its entirety from a

programme focused on product innovation to one of process innovation – the knowledge is *not* lost in the transition, but it simply may not be as useful to researchers in the other mode.[19]

Section 4.6 ventured a step beyond the standard comparative analysis of equilibrium dynamics, by investigating what happens to the trajectories in the model when the firm experiences exogenous and unexpected shocks affecting market parameters or state variables during the course of a cycle of development. It resolved the problem of 'what happens next' with numerical continuation methods. Three scenarios were considered: a rise in the interest rate, a decline in researchers' market wage, and an increase in the state of fundamental knowledge (possibly due to research discoveries in the public sector, or knowledge spill-overs from private producers in other, seemingly unrelated markets), each of which took place when applications-oriented R&D was being ramped up. Significantly, the results exhibited hysteresis effects, in that the disturbed continuation paths did not converge to the limiting levels that would have obtained had the alteration of the parameter setting occurred at the origin-point of the process. One could argue that an analysis of this kind is more informative for policy-making than is comparative equilibrium dynamics: policymakers require a mode of analysis that acknowledges that whatever the changes they might initiate, there will be short-run effects that are heavily dominated by the consequences of disturbing enterprises' ongoing investment projects. The long gestation periods and cumulative effects

[19] Apart from explaining some empirical regularities of 'new industries,' this straightforward extension of the dynamic model to a sequential setting may help to understand the important element of path-dependence that has been seen to characterise the development of certain products, in as much as the initial conditions regarding the firm's product quality and production technology will be determinant in the choice of a suitable research programme, with irreversible consequences regarding the intangible business assets it has later on. Indeed, there may be occasions in which there is a knife-edge property to the evolution, as the initial conditions to the firm may make it indifferent to alternative research programmes if the cumulative discounted profits from each yield the same figure, in which case it may randomly pick one of the two, with quite different implications for product quality, pricing, output, etcetera. See David (2000b) for a discussion of the significance of path-dependence in regards to the co-evolving technological knowledge and practice; and David (1998) for further technical details.

of knowledge investments mean that it is particularly important for dynamic analysis in the area of science and technology policy to confront these issues formally.

Having set down the results that seem most compelling in our numerical analysis, this is an appropriate juncture at which to make a commentary on the foremost limitations of the specification that has been used to get these results and the possibilities that exist for resolving these in future work. The first limitation was touched upon in Section 4.4, and relates to the allegation that the monopolist is engaging in an intertemporal variety of price discrimination by setting a progressively lower price for a good of equal quality when it is improving its production processes. As we argued there, this anticompetitive behaviour is connected to the specification of demand. A time-invariant market demand function is an imperfect representation of consumers in the context of technological change. It implies that consumers are short-sighted, acting as passive receivers of the goods they are offered by the firm, and not reacting to the technological evolution that they are witnessing, or ignoring its implications, even though they face the consequences directly in terms of the product design and the price. To take care of this shortcoming, it would be necessary to explicitly include a maximisation by consumers over the relevant time horizon; representing consumer preferences by some variant of the constant intertemporal elasticity of substitution (CIES) utility function would simplify the solution.

In respect to the production side of the model, there are at least two outstanding elements that one might wish to address in future work. The first has to do with the unmodelled fixed costs that the firms may need to finance. We say unmodelled because the total sum spent by the firm on exploratory R&D can be considered as a type of fixed cost, given that it does not directly attract a stream of revenue – the only reason why the firm does not accelerate its spending on this research is that it runs into decreasing returns to the generation of fundamental knowledge. Introducing a fixed cost into the model would shift

down the revenue curve, delaying the break-even point and driving down cumulative profits, which could dip below zero, in which case the firm simply would not operate. It would not add anything significant to the analysis unless the fixed investment had some effect on the marginal value of either type of research, or of labour in production. The optimal paths could be altered if the fixed cost was needed to fund infrastructure that was essential to the firm's research plan or to fully exploit the gains from the improved manufacturing efficiency, or was used to finance complementary investments in marketing to ensure, say, that consumers learnt about the greater quality of its products. The second element concerns the omission of an investment in durable tangible capital by the firm – in contrast to the portfolio of intangible business assets, which has been explicitly considered. This is an assumption that only is defensible on the grounds that one needs to be parsimonious in the context of an already complicated problem, as there are numerous reasons to think that these investment decisions are made conjointly.

The implicit and explicit assumptions relating to the research process constitute another area in which the specified model could be improved. The first of these assumptions, whilst important, is only mentioned in passing. It is that the firm can effortlessly finance the debt generated by its research activities. There is ample evidence that financial market imperfections and informational asymmetries place limits to the amount that companies can borrow and, in extreme cases, can force them to rely exclusively on their internal resources, in which case the cash flow generated would become central to the analysis. The second simplification of the model is of such import as to deserve a longer comment. This is that the firm estimates the return of a given time profile of exploratory- and applications-oriented research without accounting for the randomness surrounding the efforts of its researchers. There is one method that suggests itself to deal with the matter of measurable risks associated with innovation. Concretely, instead of abstractly considering the state of scientific

knowledge (S) as a modifier in the productivity of applications-oriented research, it can be postulated as the probability that applied research will result in an improved product design or production technology, whichever is applicable. In order to redefine the state of knowledge as a "probability of success function," it must evolve in such a way that it complies with the definition of a random variable at any point in time.[20] By introducing this extra structure into the model, the firm would be maximising the 'certainty equivalent' of its discounted cumulative profits. The model and the discussion made so far could then be reinterpreted in certainty equivalent terms.

Lastly, the analysis of volatility developed in this chapter can be challenged on the ground that an innovative business may have the capacity to partially foresee some of the disturbances that could impact on its profitability. This is because it could survey the potential sources of knowledge outside its direct control as well as the conditions that may change in its relevant product and labour markets. Consequently, the results of this analysis would be more relevant if the firm in question is a start-up that has successfully protected its new product through IPR, and much less so of a monopoly that has been long-established on the basis of its exclusive control of key inputs or distribution channels. One would expect the latter to anticipate and if possible to dampen the shocks in its relevant factor markets, and possibly to have a longer-term vision of the possible evolution of the economy, and its consequences on the interest rate or wages. All the same, there are probably in-built limits to this kind of protection in a context where R&D investments are irreversible and it is impossible to go 'short' in research – except by cutting back current spending. A firm would

[20] To be more precise, the equation of motion for the 'probabilistic state of knowledge' $\dot{S}$ needs to be increasing and concave in N_E to satisfy the sufficient conditions for a global maximum. Additionally, given $S(0)$ in the interval $[0,1]$, $S(t)$ must remain in the interval $[0,1]$ for any sequence of nonnegative $N_E(t)$. These conditions could be met by a variant of the logistic form or by modifying the specification employed in this chapter, for example.

be better placed to buy insurance against shocks in the interest rate, for example, for which there are financial instruments that can be bought to ameliorate the risk of fluctuations, although these would entail an extra fixed cost that needs to be recovered.

Appendices

Appendix 4.1 Reduced form differential equations and 'derived' variables

Product specification

$$\dot{\psi}_S = -\left[1-\upsilon_Z\right]\left[\frac{\upsilon_Z}{w_A}\right]^{\frac{\upsilon_Z}{1-\upsilon_Z}} \psi_Z^{\frac{1}{1-\upsilon_Z}} + \delta\psi_S$$

$$\dot{\psi}_Z = -\chi\left[\frac{\alpha[\varepsilon-1]}{w}\right]^{\frac{\alpha[\varepsilon-1]}{\alpha+\varepsilon[1-\alpha]}} \varepsilon^{\frac{-\varepsilon}{\alpha+\varepsilon[1-\alpha]}} T_0^{\frac{\varepsilon-1}{\alpha+\varepsilon[1-\alpha]}} Z^{-1+\frac{\chi}{\alpha+\varepsilon[1-\alpha]}} + \delta\psi_Z$$

$$\dot{S} = \left[\frac{\phi\psi_S}{w_E}\right]^{\frac{\phi}{1-\phi}}$$

$$\dot{Z} = \left[\frac{\upsilon_Z\psi_Z}{w_A}\right]^{\frac{\upsilon_Z}{1-\upsilon_Z}} S$$

$$L = \left[\frac{\alpha[\varepsilon-1]}{\varepsilon w}\right]^{\frac{\varepsilon}{\alpha+\varepsilon[1-\alpha]}} T_0^{\frac{\varepsilon-1}{\alpha+\varepsilon[1-\alpha]}} Z^{\frac{\chi}{\alpha+\varepsilon[1-\alpha]}}$$

$$N_A = \left[\frac{\upsilon_Z\psi_Z}{w_A}\right]^{\frac{1}{1-\upsilon_Z}} S$$

$$N_E = \left[\frac{\phi\psi_S}{w_E}\right]^{\frac{1}{1-\phi}}$$

Process specification

$$\dot{\psi}_S = -\left[1-\upsilon_T\right]\left[\frac{\upsilon_T}{w_A}\right]^{\frac{\upsilon_T}{1-\upsilon_T}} \psi_T^{\frac{1}{1-\upsilon_T}} + \delta\psi_T$$

$$\dot{\psi}_T = -\left[\frac{\alpha}{w}\right]^{\frac{\alpha[\varepsilon-1]}{\alpha+\varepsilon[1-\alpha]}}\left[\frac{\varepsilon-1}{\varepsilon}\right]^{\frac{\varepsilon}{\alpha+\varepsilon[1-\alpha]}} T^{-1+\frac{\varepsilon-1}{\alpha+\varepsilon[1-\alpha]}} Z_0^{\frac{\chi}{\alpha+\varepsilon[1-\alpha]}} + \delta\psi_T$$

$$\dot{S} = \left[\frac{\phi\psi_T}{w_E}\right]^{\frac{\phi}{1-\phi}}$$

$$\dot{T} = \left[\frac{\upsilon_T\psi_T}{w_A}\right]^{\frac{\upsilon_T}{1-\upsilon_T}} S$$

$$L = \left[\frac{\alpha[\varepsilon-1]}{\varepsilon w}\right]^{\frac{\varepsilon}{\alpha+\varepsilon[1-\alpha]}} T^{\frac{\varepsilon-1}{\alpha+\varepsilon[1-\alpha]}} Z_0^{\frac{\chi}{\alpha+\varepsilon[1-\alpha]}}$$

$$N_A = \left[\frac{\upsilon_T\psi_T}{w_A}\right]^{\frac{1}{1-\upsilon_T}} S$$

$$N_E = \left[\frac{\phi\psi_S}{w_E}\right]^{\frac{1}{1-\phi}}$$

Appendix 4.2 An example of Matlab's bvp4c solver for the model of product innovation

```
function results = product

% Benchmark parameters: rho=0; x=.75; e=1.25; a=0.75; v=0.5; phi=0.75;
% w=1; wa=2; we=2; t0=100;
rho=0; x=.75; e=1.25; a=0.75; v=0.5; phi=0.75; w=1; wa=2; we=2; t0=100;

% Initial guess for the boundary value problem
solinit = bvpinit(linspace(0, 25), @productinit);

% Boundary value problem solver for the differential equations in "productode,"
% boundary conditions in productbc and initial guess structure solinit
sol = bvp4c(@productode, @productbc, solinit);

% Numerical evaluation of the solution in a specified range
tint = linspace(0,25,50);
Stint = bvpval(sol,tint);

% Derivation of control and other relevant variables
l = ((a*e-a)/(e*w))^(e/(a+e-a*e)) * t0^((e-1)/(a+e-a*e)) * Stint(4,:).^(x/(a+e-a*e));
na = (v/wa)^(1/(1-v)) * Stint(1,:).^(1/(1-v)) .* Stint(3,:);
ne = (phi/we)^(1/(1-phi)) * Stint(2,:).^(1/(1-phi));
price = t0^(-1/e) * Stint(4,:).^(x/e) .* l.^(-a/e);
quantity = t0 * l.^a;
revenue = price .* quantity;
wage_bill = w*l + wa*na + we*ne;
profits = revenue - wage_bill;
```

```
% Output from the 'product' function
results = [Stint' l' na' ne' price' quantity' revenue' wage_bill' profits'];

% Function that evaluates the differential equations for the co-state variables psi-s and psi-z,
% and the state variables s and z
function dydt = productode(t,y);
rho=0; x=.75; e=1.25; a=0.75; v=0.5; phi=0.75; w=1; wa=2; we=2; t0=100;
dydt = [  rho * y(1) - x * ((a*e-a)/w)^((a*e-a)/(a+e-a*e)) * e^(-e/(a+e-a*e)) * ...
          t0^((e-1)/(a+e-a*e)) * y(4)^(-1+x/(a+e-a*e))
          rho * y(2) - (1-v) * (v/wa)^(v/(1-v)) * y(1)^(1/(1-v))
         (phi/we)^(phi/(1-phi)) * y(2)^(phi/(1-phi))
         (v/wa)^(v/(1-v)) * y(1)^(v/(1-v)) * y(3) ];

% Function that computes the residual in the boundary conditions -- the two initial conditions
% for the state of knowledge and technology, and the boundary (transversality) conditions
% for the co-state variables
function res = productbc(ya,yb)
res = [ ya(3) - 100
        ya(4) - 100
        yb(1)
        yb(2) ];

% Function used for the guess structure of the co-state and state variables
function guess = productinit(t)
guess = [ 0
          0
        100
        100];
```

Appendix 4.3

Figure A4.1 Process innovation with and without exploratory R&D

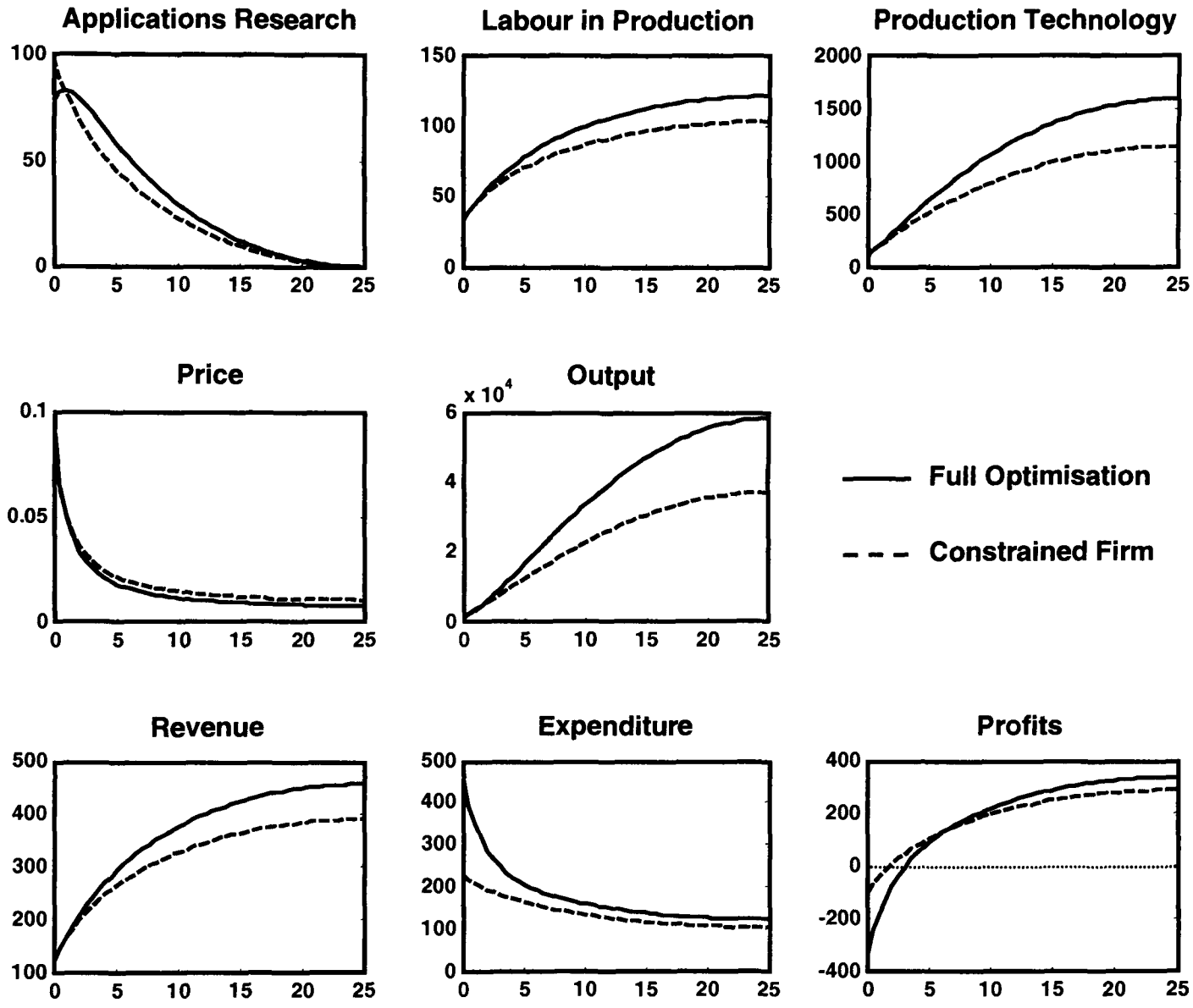
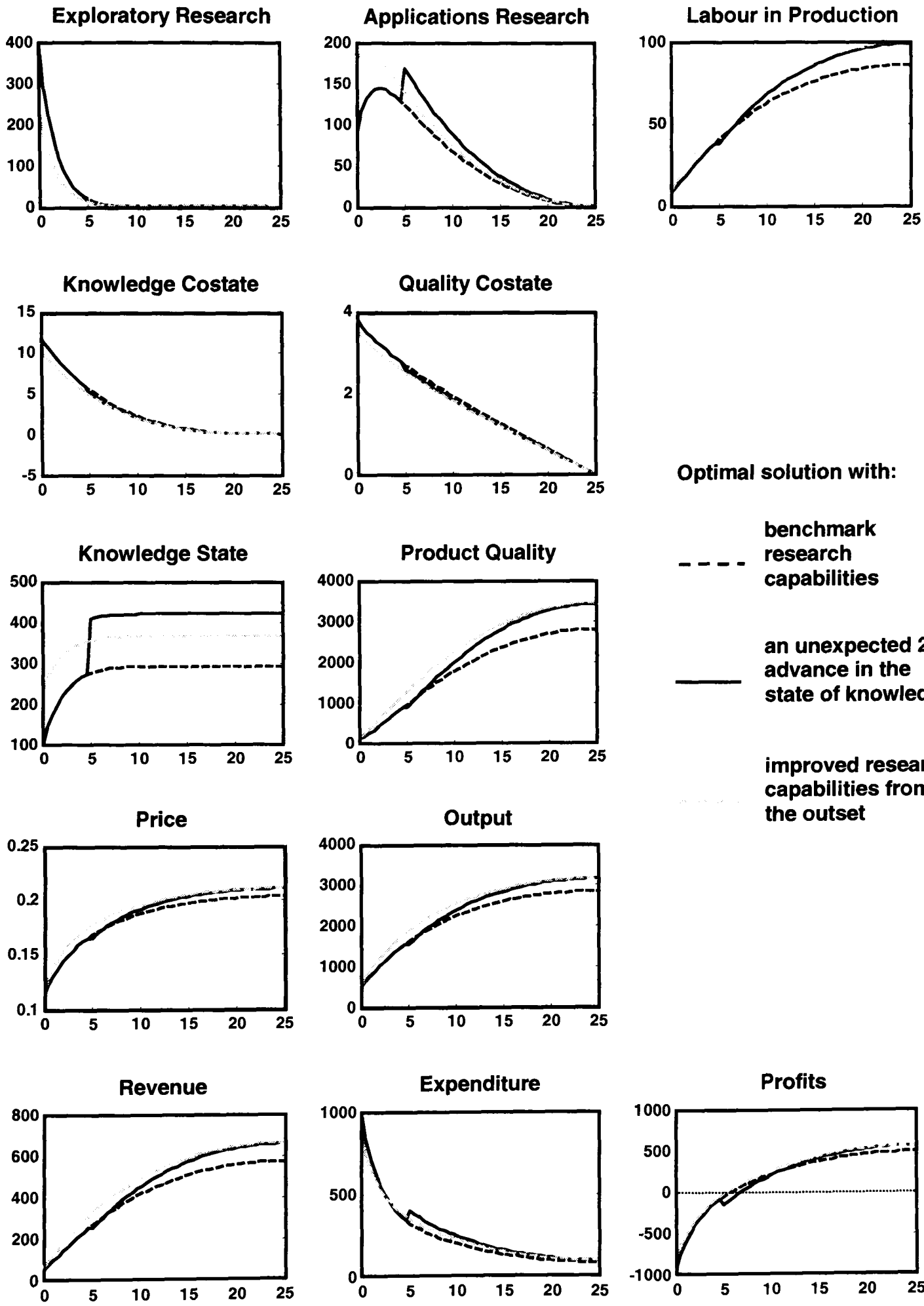


Figure A4.2 Optimal response to an unanticipated advance in the state of knowledge



Chapter 5

Managing R&D and production across multiple product lines

5.1 Introduction

The last two chapters investigated the research and production plans decided upon by a monopoly enterprise seeking to maximise its discounted stream of profits in a single cycle of development, the time of which was fixed in advance. This chapter extends this theoretical framework by embedding the model in a multiple product line context, where there are supply-side and demand-side externalities across the product lines. To simplify matters, the analysis will assume that the product lines are marketed one at a time, each line completely superseding earlier ones. The product lines of software and automobile companies, in which only the latest in the successive generations is usually marketed, seem to be plausible examples of this sort of development process.[1] The former is doubtless the better of the two, as the models presented here are all worked out when monopoly conditions are prevalent. Although this chapter is concerned with the case of product innovation, the framework introduced and its results could be rearticulated in terms of interrelated families of improving production processes and a fixed product design by recourse to the solution for process innovation obtained in previous chapters.

[1] According to a member of the Software Engineering Institute, for software development to classify as following a product line approach, the set of software-intensive systems must "share a common, managed feature set satisfying a particular market segment's needs or mission and that are developed from a common set of *core assets* in a prescribed way" Northrop (2002, p. 32). Her list of core assets includes: "the architecture, reusable software components, domain models, requirements statements, documentation and specifications, performance models, schedules, budgets, test plans, test cases, work plans, and process descriptions." Countless examples can be provided that satisfy this definition, such as Microsoft's Windows or Word, Adobe's Acrobat Reader, Aladin Enterprises' Ghostscript, or Netscape Communications Corporation's Navigator. What they also have in common is that only the newest product version is generally offered and that the companies enjoyed a great deal of market power in their respective product markets for a period of time.

Firms employ a variety of project evaluation methods to settle the most efficient allocation of resources across several R&D projects. In those cases where multiple projects are to be undertaken concurrently, the aim of these methods is generally to provide a ranking of the projects. Osawa (2003) briefly reviews ranking methods such as the NewProd System or Decision Analysis before evaluating a scoring method of his own invention, the NS-method, which rates projects based on the following evaluation criteria: strategic importance and technological effect, probability of realization, sales, profits, and R&D efficiency (in profit terms). A weakness shared by these evaluation tools is that, by focusing on the possibilities for discriminating among projects, they fail to deal with the correlations and commonalities between them in terms of either development or marketing.[2]

The literature explicitly dealing with a *connected succession* of product lines is much sparser. Moreover, theoretical models in this area tend to assume a functional relationship between revenue and the time of introduction of each product generation, rather than deriving this relationship from a firm's investment decisions. This simplification is made by Zeckhauser (1968), whose approach is commented on later in the chapter, and also by Mahajan and Muller (1996). The latter employ a standard diffusion process to model the 'flow dynamics' in each product generation, and go on to find the optimal timing of successive generations for a monopolist, which turns out to be a "now or at maturity" rule, in which the new generation is introduced as soon as it is available or until the previous

[2] But maybe some lessons about the "optimal R&D project portfolio" can be extracted from theoretical examinations into the efficient scheduling of limited production capacity when there are a variety of products from a known "book of blueprints," that could be manufactured. For instance, in a numerical analysis of an integer programming problem, De Matta and Guignard (1994) find that costly product changeovers at the production line-level often lead to an optimal "capacity-oriented scheduling" in which some but not all of the production lines are specialised – i.e., they manufacture only one product throughout the relevant time span. This could suggest that some research teams should be allocated to a single R&D project, so long as this has a sufficiently high anticipated demand and probability of success; but other teams should be made to work on several projects at a time, rather than eliminating the projects altogether, which is what the ranking method would suggest. This class of multi-period capacity production planning problems, sometimes called line segmentation problems, have been studied through various programming techniques by Bitran, Haas and Matsuo (1986), Ahmadi and Matsuo (1991), Balakrishnan and Vanderbeck (1999), among others.

generation has reached maturity. They then apply their results to data from four generations of IBM mainframe computers, which leads them to suggest that "their time to market should have been shorter" in the cases of the 360 and 370.[3]

Central to the vision of the innovation process in this chapter is the notion that fundamental knowledge and product improvements generated by the firm's R&D spending are useful to the next sequences of development in different but related product lines. Research tools, once acquired, can accelerate subsequent research efforts, and the new skills learnt by researchers as well as any intangible but disembodied knowledge assets permanently improve the firm's research capabilities as it moves from producing and improving, say, in the case of Microsoft, Windows ME to Windows XP. Some components and design features become integrated to the overarching 'product concept' as time advances; that is, the "practical knowledge" embodied in the product design is transmitted or, to use the standard terminology, spills-over to new product lines. But spillovers will generally not be complete, as it is unlikely that these knowledge assets are equally useful to follow-on products. And there may be sunk costs involved with "switching" having to do with the reconfiguration of the production technology, so that it can be efficiently redirected from the production of one product line to the next, or the marketing expenses informing customers about the novel aspects of newer product generations on offer.

To model these cumulative externalities in a simple way, we introduce a set of parameters specifying the degree of spillovers (in knowledge, quality and technical efficiency) as well as any gains to developing new product lines in terms of increasing market size. These parameters will govern the transitions in a sequence of optimal control problems,

[3] Their work builds on that of Wilson and Norton (1989), who had found a "now or never" rule to be optimal with a similar theoretical framework. Note that Vickers' (1986) study of sequential patent races or Delbono's (1989) extension should not be included in this category because they are exclusively concerned with the relationship between market leadership, past profits and current incentives to innovate. They do not seek to model the increased development possibilities that may follow from, e.g., accumulating patents in a specific set of products.

'joining' together the evolution in different product generations – each represented by the dynamic model set out for one product lifecycle. By framing the problem in this manner, a numerical solution can readily be obtained using a 'continuation method' in conjunction with the two-point boundary value problem algorithm that was employed in the last chapter. Geometric "envelope curves" are then used to typify the overall trends in the optimised time paths for control, state, costate and composite variables (such as revenue or profits) across the product lines. Envelope curves were central to the "technological forecasting" literature's extrapolation of the innovation potential of product lines. One of the aims of this chapter is to account for the empirical observations made in that literature with a modelling approach having more rigorous microeconomic foundations. Section 5.2 gives a detailed account of the model's notation, assumptions and methodology.

The research questions asked are: What role could various classes of spillovers play in the optimal growth paths in performance, output, productivity and profitability, as well as for the long-term viability of the research programmes that will be seen to prop up the geometric envelope traced over these paths? In particular, under what conditions might the bounds on growth imposed by decreasing returns in the R&D process and production be overcome? And does the co-evolution in prices, output, revenue, and other variables replicate that of an individual product lifecycle, or do other patterns surface with multiple cycles of product development? These questions are addressed in three different settings.

Section 5.3 starts off with a simple set-up, in which each product line is marketed for a predetermined and equal amount of time, in order to give a general indication of the feasible envelope curves under a wide range of parameter values. The simulation results presented there indicate that the envelopes to the progression in performance, output, productivity and profitability have a positive slope. The envelopes are more likely to be convex (i.e., to be characterised by increasing growth rates) if: knowledge spillovers are

plentiful; the production technology is equally efficient in the manufacture of any product variety; the market is more responsive to the introduction of new product lines; and the discount rate is low or zero. But if the parameters are biased in the opposite directions, the solution to the sequential continuation problem generates a concave evolution for the control, costate and state variables across the envelope, resembling that of a single product lifecycle.

The process of growth is sustained by the traverse across an ordered sequence of product lifecycles, each with an S-shaped evolution. Individual product lines run into decreasing returns to R&D investments, limiting the optimal research intensities and growth dynamics over each lifecycle. But, thanks to knowledge spillovers and market expansion, the trend in aggregate R&D spending may be stable or increasing over time, thus supporting positive and accelerating rates of change in revenue, productivity and profitability across the product lines. The results obtained from this analysis may be interpreted as having a bearing upon long-run aggregate economic growth – following Kuznets' view of the aggregate path as reflecting the expansion and decline of particular industries that build upon each other through demand-side or supply-side externalities – in contrast to the movement along a steady-state path typical of new growth theories.

The next two sections aim to capture various goals that firms seem to have in mind when devising a plan for their R&D and output in multiple product lines with knowledge spillovers but costly switching. In Section 5.4, the firm seeks to maximise its discounted profits by selecting the *number* of product lines to develop and produce during a given time span. Alternatively, Section 5.5 determines the potential for "accelerated product development" when this is restricted to be financially feasible, in the sense of allowing the firm to regain all of its sunk costs. With respect to the 'optimality' of the switching decision, the results suggest that there is a trade-off between developing one product line extensively vs. marketing a string of product lines in quick succession; depending on the pecuniary cost

of the transitions and the magnitude of the externalities, this may lead to an interior solution (with two or more product lines) or one with a single product lifecycle. A financially-feasible research programme calls for a long planning horizon for the first product line, characterised by heavy investment in exploratory research, and then much shorter lifecycles for the ensuing product generations, with less R&D spending, and that mostly focused on applications-oriented research. The last section discusses the empirical consistency of the results and makes an assessment of the limitations of this numerical analysis.

5.2 Model and methodology

5.2.1 Theoretical framework

Consider a firm that will market a sequence of M science-based product lines or product families,[4] a collection of product lines that build upon each other in terms of the design qualities and other attributes demanded by consumers. The interrelation between the product lines via the market demand function is but one of the elements they share. The product families are derived from a common 'platform,' that is to say, the development possibilities of each of the M product lines depend on the strengths and weaknesses of a set of core knowledge assets – among which one may include the overarching 'product architecture,' documentation and specifications, test plans and cases, and other items. Product lines are related at the level of production as well, although the compatibility or congruence of the production system to different lines may be less than perfect; switching production across

[4] In this paper, the two terms will be used interchangeably, in accordance with the majority of the literature. But note that the metaphor of a 'product family' could be thought to imply a broader conception, including: more distant "kin" such as spin-offs from the research tools and products developed by the firm; concurrent marketing of "siblings" aimed towards different segments of the market; products devoid of "legitimacy," in that they are developed with research tools not produced in-house; or "adopted" products, bought from independent inventors or other companies. None of these "relatives" are the subject of the present study.

products may disrupt the manufacturing processes, with negative consequences for labour productivity.

To simplify the analysis, and allow this extension to make full use of the numerical methods and results derived in the preceding chapters, we assume that the product lines are developed and marketed one at a time. The production and research controls are selected by maximising the discounted stream of profits from *each* product line, conditional on the evolving state of the product design's quality, the efficiency of the production technology, and the fundamental knowledge available to the firm's researchers. Thanks to these assumptions, the basic dynamic structure and the set of assumptions outlined for a single product line can be embedded in a sequential structure of interrelated product development. The timing of the ordered succession of M product lines will be given by the series $\{\tau_1, \tau_2, ..., \tau_M\}$, where $\tau_i > 0$ is the time during which the i-th line is improved and marketed. To illustrate the nature of the solution in the numerical continuation problem to which this dynamic structure leads, it is initially assumed that the planning horizon for each of the M product lines is the same. This restriction is relaxed in Sections 5.4 and 5.5 to investigate different criteria that may shape the timing of the firm's switching decision, and the implications this holds for exploratory and applications-oriented R&D activities.

The cornerstone of this analysis is the specified model of Chapter 4, of which we now present a highly summarised account, and then go on to describe the additional structure and assumptions needed in a multiple product line context. A firm starts off with a product of quality $Z(0) > 0$, which it alone can manufacture. By employing $L(t) \geq 0$ labour inputs, it produces output $q(t) = T(t)[L(t)]^{\alpha}$, $\alpha \in (0,1)$, where $T(t) > 0$ is the state of the production technology at time t; initially $T(0) > 0$. The inverse market demand function $p(t) = [Z(t)]^{\chi/\varepsilon}[q(t)]^{-1/\varepsilon}$ has price elasticity of demand $\varepsilon > 0$ and elasticity of demand with

respect to quality $\chi > 0$.[5] The firm can improve its product design by employing research workers. If the number of applications-improving researchers is $N_A(t) \geq 0$, there is a (certain) gain in quality $\dot{Z}(t) = [N_A(t)]^{\nu_Z} [S(t)]^{1-\nu_Z}$, $\upsilon_Z \in (0,1)$, where $S(t)$ is the state of fundamental knowledge. Initially, the firm is endowed with a knowledge base of measure $S(0) > 0$, but exploratory research activities can augment this: if $N_E(t) \geq 0$ researchers are employed in this task, the firm gains $\dot{S} = N_E(t)^{\phi}$, $\phi \in (0,1)$. The supply of labour, whether for production or research, is assumed to be perfectly elastic; the market wage of workers in production is denoted as $\omega > 0$, whereas the wages of research workers in exploratory and applications research are $\omega_E > 0$ and $\omega_A > 0$ respectively.

Since the firm's optimisation problem will be divided into M optimal control problems corresponding to successive product lines, it is convenient to introduce some additional notation to keep track of the evolution of the state variables in each cycle of product innovation. The quality of the i-th product line in the sequence when it has been marketed and improved for t_i periods is denoted as $Z_i(t_i)$. Similarly, let $T_i(t_i)$ and $S_i(t_i)$ denote the states of the production technology and fundamental knowledge, respectively, when line $i = 1,\ldots,M$ has been developed for t_i periods. For example, $Z_3(5)$ would stand for the quality attained by the third product line in the series when it has been in development for five periods. Suppose that the first two lines had been marketed for 25 periods. Then the state of the product's quality at that moment would be denoted as $Z(30)$; similarly, the state of knowledge at that point could be denoted as $S_3(5)$ or $S(30)$.

[5] The admissible parameter subspace needs to be further reduced if the problem's sufficiency conditions are to be met. In particular, concavity of the objective functional requires the elasticities in the functional relationships specified in the text to lie within a narrower range. The limits are detailed in the second section of Chapter 4.

There are knowledge externalities in the transition from the i-th product line to the $i+1$-th. The 'degree of spillovers' in fundamental knowledge, product design, and production technology is regulated by parameters μ_S, μ_Z, and μ_T respectively. These parameters are assumed to be constant, each in the closed interval $[0,1]$ and the same for any of the $M-1$ switches between product lines – the latter assumption is made because there is no *a priori* reason or reliable empirical evidence for selecting a particular sequence of transitions (i.e., one that is monotone increasing, or concave, or otherwise). Therefore, at the time when the firm switches to the $i+1$-th line, its state of fundamental knowledge will equal $S_{i+1}(0)=\mu_S S_i^*(\tau_i)$, where $S_i^*(\tau_i)$ is the optimised value of the state variable at the end of the i-th lifecycle. The transition equations for the states of the product's quality and production technology are $Z_{i+1}(0)=\mu_Z Z_i^*(\tau_i)$ and $T_{i+1}(0)=\mu_T T_i^*(\tau_i)$. When the parameters take values that are strictly less than 1, there will be partial or incomplete knowledge spillovers; these result in downward jumps in the state variables at product line transitions.[6]

In addition to the fundamental and practical knowledge carried over, the firm may benefit from demand-side externalities because the number of potential customers or their income may be growing over time, or newer product families may have some distinguishing attribute in the eyes of consumers; the latter would have to be thought of as independent or 'orthogonal' to the measure of quality $Z(t)$ that the firm is purposefully improving. There are many ways of modelling this drift. That undertaken in this chapter assumes an expanding market as the firm introduces newer product generations. Let C_i denote the size of the market for the i-th product line. Then, we assume that the market for the $i+1$-th line will equal

[6] One can provide many different rationales for the presence of partial spillovers. For instance, the firm's fundamental knowledge base may not be as applicable to the design-improving research efforts in newer product lines ($\mu_S<1$). Or there could be efficiency losses as the firm shifts its production from one product line to the next if it had invested in a manufacturing system that was highly efficient in the production of its first innovation but not very flexible ($\mu_T<1$).

$C_{i+1} = C_i + \mu_C$, where $\mu_C \geq 0$ is a fixed parameter measuring the expansion in successive product lifecycles. This specification implies that the market size is a function of the number of product lines that have been marketed: if C_1 is the market for the first line, the $i+1$-th will have a market size $C_{i+1} = C_1 + i\mu_C$.

The transition across successive product generations considers a third element. This is a fixed switching cost $F_S \geq 0$ that needs to be paid by the firm when it starts a new cycle of product development. A variety of fixed outlays of this kind have been identified in the management of innovation literature. These include one-time machine set-up costs, as machinery (even in reasonably flexible production systems) typically needs to be cleaned, re-tooled, and rectified. Firms may also have to pay for overtime labour to resume production as fast as possible if changeovers disrupt the normal functioning of manufacturing processes.[7]

The optimisation problem is different in each of the three extensions with which we study the dynamic management of multiple product lines. The common denominator of all three is that the time path for the control variables is selected to maximise the discounted sum of profits accruing from each of the ordered succession of M interrelated product families. The discount rate is $\delta \geq 0$. In other words, the optimal time paths for the control variables are determined as the maximum *within* each cycle of product innovation, conditional on the evolving frontier of fundamental knowledge, product quality and production technology. As we explain below, the numerical continuation method introduced in Chapter 4 can be employed to solve this problem. There is another aspect to the solution related to the time structure, which is exogenous in Section 5.3, but endogenised in the next two sections.

Specifically, the analysis undertaken in Section 5.3 assumes that the timing is exogenous, each product line being improved and sold for an identical amount of time.

[7] In tile manufacturing companies, for example, adjustments at the "moulding section" (e.g., washing of machines and changing dies) are usually performed outside regular work hours to prevent the "curing" process from being interrupted; see De Matta and Guignard (1994).

Instead, Section 5.4 seeks to determine the number of product generations that maximise net profits (that is, profits minus any fixed switching costs) if the total development time for the product families is preset. To this end, it is necessary to calculate and compare net profits when there are a different number of product lines. Another way of thinking about this problem is that the firm is trying to find the optimal time for its individual cycles of product development under the restriction that each should last for the same length of time. Lastly, Section 5.5 estimates the minimum period during which successive product lines must be marketed if *each* is to break-even; in other words, the time of development of every product generation such that its own profits pay for the cost of switching to the next one. In this case, the duration of each cycle is variable. The specific assumptions and solution procedures in these problems are presented in the relevant sections.

5.2.2 Numerical continuation methods

Chapter 4 showed that the optimal controls for the case of a single product lifecycle could be solved numerically using Matlab's solver *bvp4c*, which is designed to tackle two-point value boundary value problems (BVP's), including those that are non-linear. The solver required three types of inputs: the set of differential equations that describe the system's evolution, which are derived from the necessary conditions for a maximum in the optimal control problem; boundary conditions at the initial and terminal times; and an initial guess of the solution.[8] Later in that chapter, we examined the adjustment process of a firm facing environmental volatility with a numerical "continuation method." Volatility was defined as an unforeseen variation in the structural conditions, such as the interest rate or wages, some time into its pre-optimised production and research programme. The method consisted of a

[8] The reason for making a guess is that BVP's, unlike initial value problems, can have multiple solutions; our problem was extremely robust to the initial guess, and to a variety of other 'sensitivity tests,' all of which are discussed at length in Section 4.3.3 of Chapter 4.

solution to two optimal control problems: the original problem of the firm; and another for the subinterval going from the time of the shock, which was denoted t', to the terminal time t_1, taking into account of the new parameter values and any adjustments to the state variables.

When a change to a parameter such as δ or w_A was examined, the initial boundary conditions corresponding to the 'continuation problem' were the optimised values of the state variables computed for the original problem at t'; $S^*(t')$ and, depending on the mode of innovation, either $Z^*(t')$ or $T^*(t')$. A new simulation could then be run for the interval $[t', t_1]$ with the new parameters, but the original transversality conditions, which compelled the two current-value costate variables to have a value of zero at t_1. To model a discrete change in the state of knowledge of 'size' Δ, the initial value of $S(t)$ in the continuation problem was $S(t') = S^*(t') + \Delta$. The other state variables had a starting value of either $Z^*(t')$ or $T^*(t')$, whereas the parameters and transversality conditions were left as in the original problem. These discrete and unanticipated variations in the parameters or the state variables did not jeopardise the existence of a solution for the 'continuation problem,' which was simply a (shorter) duplicate of the problem for the entire planning horizon, and satisfied all of the necessary and sufficient conditions for a maximum.

The same method will be employed to study the problem at hand, but in the context of multiple product lines the process must be repeated as many times as there are switches between lines; e.g., Section 5.3 assumes that there are five product generations ($M = 5$), which means that four continuation problems need to be solved. Depending on the assumptions regarding the transition parameters, each subsequent problem has different parameter values and initial conditions for the state variables. In the "benchmark case" we introduce, the parameter corresponding to the market size increases by .1 in each subsequent control problem ($\mu_C = .1$); the optimised values for the states of fundamental knowledge,

product design and production technology determined at the terminal time of one problem are used as initial conditions for the next one ($\mu_S = \mu_Z = \mu_T = 1$). With partial spillovers, the end values for the state variables need to be adjusted to get the initial values for the next problem using the transition equations $S_{i+1}(0) = \mu_S S_i^*(\tau_i)$, $Z_{i+1}(0) = \mu_Z Z_i^*(\tau_i)$, and $T_{i+1}(0) = \mu_T T_i^*(\tau_i)$.

It is important to make this method's strengths and weaknesses clear from the outset. It can generate a numerical solution that is optimal within each product's development cycle and which takes into account of the changes to the firm's production and research capabilities resulting from its ongoing investment in R&D and exogenous supply- or demand-side changes. It is not the "fully optimal" solution to the optimisation problem, as the effects of these investments further into the future (beyond each optimised product development cycle) are not included in the objective functional maximised by the firm. In technical terms, the transversality conditions to each continuation problem are the same as in the single product cycle model, and as a result R&D tends to zero as each product line approaches its terminal horizon τ_i. This limitation of the solution method is what prompts us to qualify the knowledge spillovers and demand-side changes across product lines as *externalities*.[9]

The presence of these intra-firm externalities can be accounted for in several ways. One is by supposing that the firm marketing the sequence of product lines is myopic. Its research division ignores the complementarities across consecutive product lines when working on any given one. This could be a result of the goals of the R&D projects. Section 5.5 explicitly analyses an R&D programme for "accelerated product development," where the firm's goal is to launch a new product line as soon as the one in progress breaks-even. Alternatively, the firm's management may fail to appreciate the strength of the connections

[9] To the best of our knowledge, the numerical research tools that are commercially available at this time cannot solve the class of optimal control problems having endogenous 'jumps' in the state variables; a numerical algorithm would have to be expressly developed for the task. Alternatively, the last section of the chapter discusses changes to the model's set-up to generate a solution closer to the fully optimal one.

between one product line and the next. Investment and production plans for successive product lines also could become uncoupled if different teams are responsible for each line, as this may give rise to problems of communication and coordination. Another, quite different explanation for the external nature of the supply and demand-side spillovers is that a different firm is developing each product line. This could be the case if firms successively obtained patents on innovations opening new product lines or on radical innovations that render previous patents obsolete. In this interpretation, the model would have successive monopolists replacing each other. It is natural to suppose that each selects a set of optimal controls by maximising its own profits.[10]

5.2.3 Geometric envelope analysis

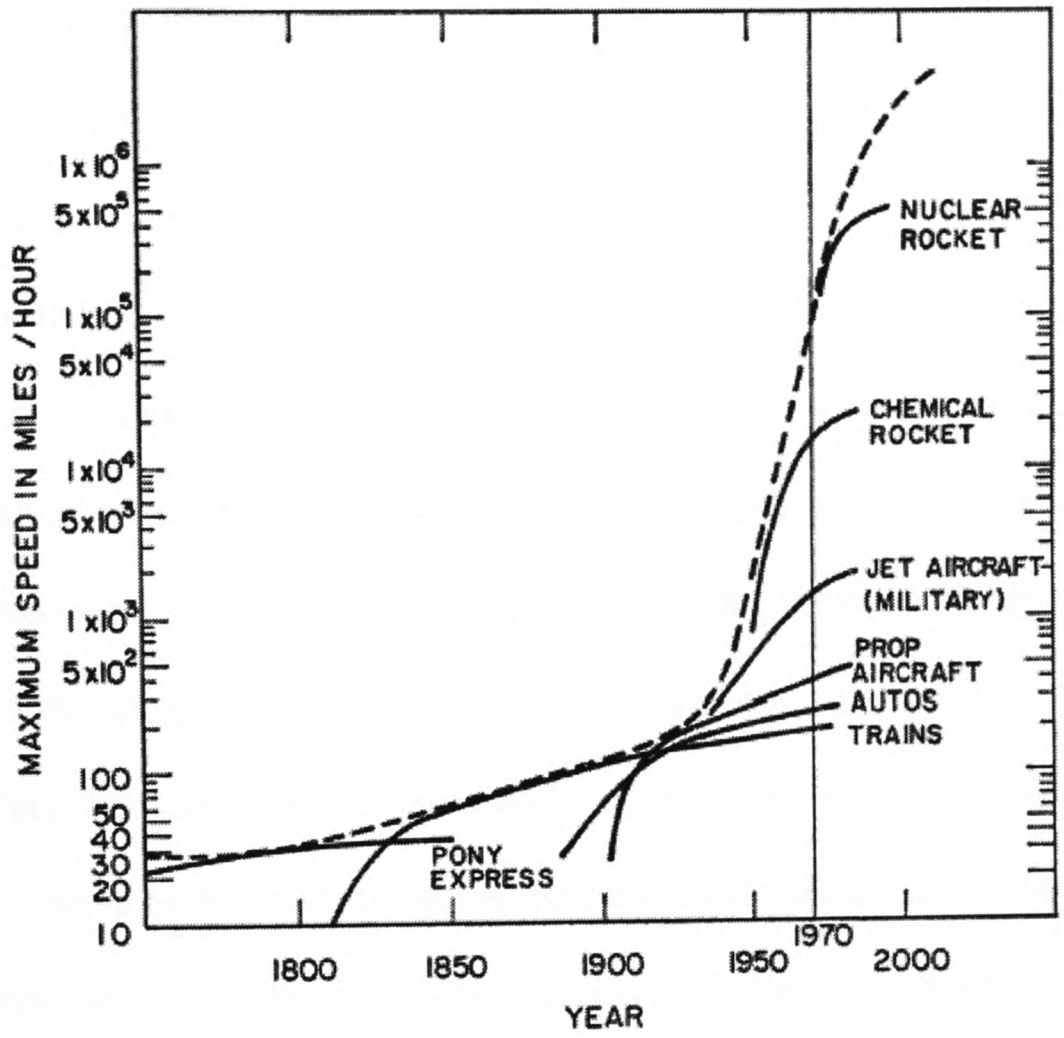
An "envelope curve" is one that encloses or envelops a series of (possibly overlapping) curves, and is tangent to all of them at at least one point.[11] In the industrial economics literature, an early application of this geometrical construct is present in Chamberlin's (1948) analysis of the long-run cost conditions faced by a firm that can select among a variety of plants, each with a different average cost curve. In his analysis, the envelope below the short-run average cost curves was U-shaped as well but was broader than the curves for individual plants. Other examples, old and new, can be found in different areas of economics; for

[10] This structure would resemble that in Aghion and Howitt (1992), but there are significant differences. Without explicitly including a stochastic process governing the arrival of these innovations, our model cannot determine the form taken by the "creative destruction" effect. Instead of a per-period reduction in R&D spending, this would consist of a change in the time profiles of the heterogeneous R&D investments and output. Extending the model in this direction calls for careful thought about the relationship between the R&D activities we are considering and the arrival rate of radical innovations, and more significantly about the R&D expenditures made by *entrants*. This is a modelling exercise we shall have to leave for further work. I would like to thank Prof. Edward Steinmueller for suggesting this interpretation of the model's structure.

[11] For any given problem, there may be innumerably many envelope curves that satisfy the "tangentiality" property – or none, if there are kinks in the individual curves over which it is drawn or one curve is completely above the others. As a result, the envelope, when it exists, is generally taken to be the 'straightest' curve that is tangent to all the individual curves in question; i.e., the curve with the least number of turning points, corresponding to the polynomial function of lowest order.

example, Baldwin's (1948) research into offer curves in international trade theory, Blitz's (1958) study of the optimal longevity of capital equipment, or Bitros and Panas's (1988) estimates of the bid functions for passenger car characteristics in Greece.

Figure 5.1 The evolution of performance in different transport modalities[12]



The usefulness of an envelope analysis in the context of technological change was pointed out by the "technological forecasting" literature, which sought to provide more accurate forecasts about the future evolution of technological attributes and parameters.[13] Robert Ayres, one of the pioneers of this approach, explained how envelope curves could be used to extrapolate the long-run trends in the innovation potential with multiple product lines. Figure

[12] Source: Originally in Ayres (1966), and later reprinted in Gerstenfeld (1991, p. 12).
[13] Other methods did not require envelopes. Some were based on simple trend extrapolation, others on the opinions of committees (the Delphi procedure), still others on morphological analysis, correlation models and simulations. Joseph Martino (1986) has written a textbook summarising the state of the art of technological forecasting. See also Bright (1968), a compilation of survey articles written by the precursors in this literature.

5.1 reproduces one of the figures provided by Ayres (1968), which depicts the maximum speed attained by different transport modes. The evolution in each technology's performance and the envelope across product families, in this case very broadly defined, is characteristic of the trends observed in many industries. Its characteristic features are that "each new type of machine takes the lead for a short period and then, because of inherent limitations, it reaches the end of its phase of rapid improvement while a newer invention escalates to a higher level. The "envelope" is a curve which approximates the general trend and is tangent to the individual performance trends" (Ayres, 1968, p. 78).

Thanks to their background in physical and engineering sciences, researchers like Ayres had a very clear understanding of the technical basis for these projections, as well as of the limits of such extrapolations into the more distant future.[14] They did not go as far as to construct a theoretical model that could accommodate the empirically observed dynamics of technological innovation. Nor has the mainstream economic literature had much need for this type of analysis, because attention is usually centred on steady-state equilibrium paths – clearly, the envelope of this sort of path is the path itself, as there are no transitions between different technological trajectories to take into consideration. This chapter adopts an envelope approach to describe the evolving upper bound to labour inputs into manufacturing, price, performance, output, revenue, productivity and profits as the firm migrates it R&D and production activities across interdependent product generations, and to determine the conditions under which these variables exhibit constant or increasing growth rates. This heuristic device is not applied in the case of R&D expenditures because the transversality conditions generate time paths that have sharp peaks at the outset and fall down to zero after

[14] Notice that new product lines do not initially dominate the established ones in terms of performance. In the example of transport modalities, the first commercially available trains were not as fast as horse-driven carriages, nor were the automobiles of the late 1800's as fast as trains of that period. Even when performance was exceeded in terms of speed, older product generations still maintained an advantage in other dimensions, such as safety, or fuel costs per mile at the prices then prevalent. These aspects are not modelled in this chapter.

each cycle of product development, and there is therefore no curve that can be tangent to all of the component curves across the extended evolution.

5.3 Development of multiple product lines with exogenously given transitions

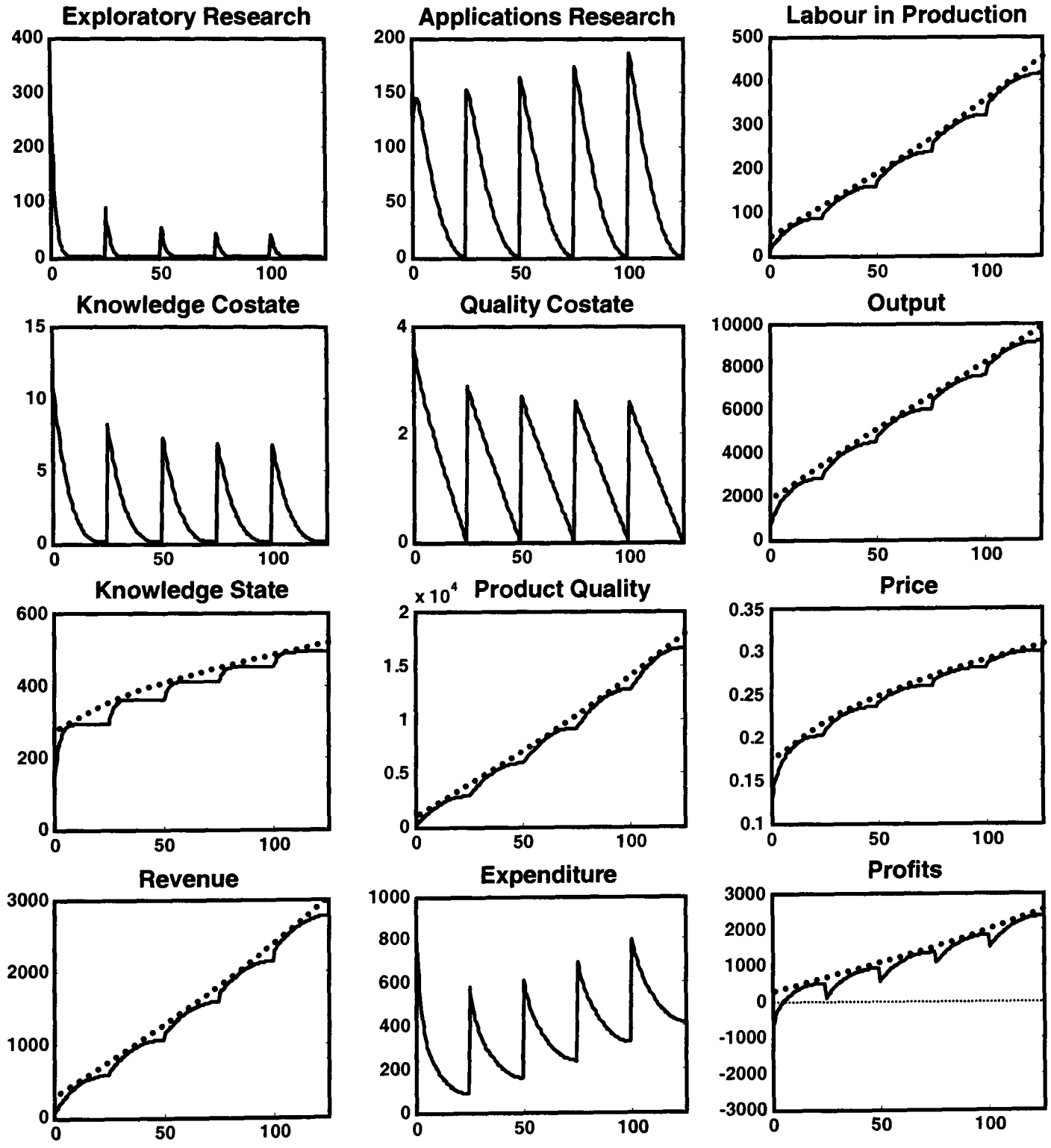
This section investigates the properties of the problem's solution set in a simple specification where five product lines ($M = 5$) are marketed for an equal duration of time; $\tau_i = 25$ for $i = 1,...,5$.[15] Initially, the parameters that govern the evolution of the system within each cycle of product development are left as in the 'benchmark specification' of Chapter 4: ε=1.25, χ=.75, α=.75, υ_Z=.5, ϕ=.75, w=1, w_A=2, w_E=2, δ=0, $Z(0)$=100, $T(0)$=100, and $S(0)$=100. Additionally, it is assumed that: there are complete spillovers of knowledge ($\mu_S = 1$), there is no loss in production efficiency at the transitions ($\mu_T = 1$); the quality of a product line when it is initially marketed exactly equals that of the line that is given up ($\mu_Z = 1$); the size of the market grows with the introduction of each product generation ($\mu_C = .1$); and, there is no fixed switching cost ($F_S = 0$). Other scenarios are studied through comparative equilibrium dynamics, the results of which are presented later in the section.

Figure 5.2 maps the results from the simulation with these parameter values. It shows a cyclical pattern of R&D, in which spending within each product line follows the path identified in Chapter 4. The solution is characterised by an initially high but declining rate of investment in exploratory research, and a subsequently rising and then declining rate of application-oriented R&D. This is only natural because the research programme maximises

[15] The situation where the transitions between product lines are predetermined and occur at equal intervals of time does not pretend to be reflective of a concrete case of innovation, but there may be factors pushing companies towards a periodicity of this sort. For instance, the firm may want to establish a long-term plan with definite dates for its technological projects to secure external financing. The decision process may also be structured in this way to have the release dates for new product lines coincide with events that heighten their visibility, thereby increasing sales; e.g., new automobile models are introduced at the annual North American International Auto Show in Detroit.

profits within each product line, not for the entire horizon. What is revealing is the trend in exploratory- and applications-oriented research, which is declining for the former mode of R&D but increasing in the latter.

Figure 5.2 Research and production across multiple product families

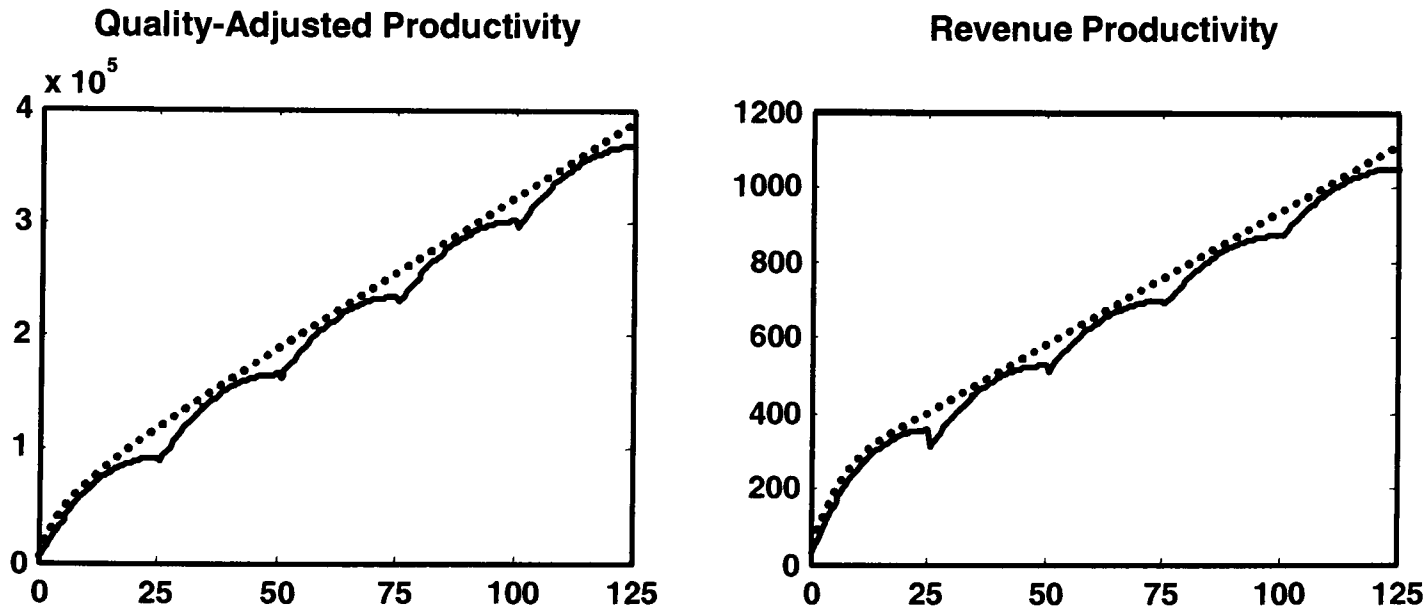


These controls result in a steepening progression in product quality over time; the envelope curve drawn above the individual product line's time paths denotes the accelerating improvement of the product design. This is a clear break with the results for a single cycle of

product development, where decreasing returns in research and production, in conjunction with a fixed planning horizon, led to an eventual decline in research for applications and a consequent slowdown in the rate of improvement of the product's performance. With this set of parameter values, the state of fundamental knowledge increases, but the 'steps' are smaller in successive product lifecycles.

Turning to the overall trend in prices, this can be seen to follow an increasing path, but with smaller increases as the firm moves from one product line to the next, replicating the concave path in the individual product lines. The solved path for labour inputs to production is similar to product quality: it has an envelope that is increasing at a constant rate or even accelerating slightly. The envelope to the periodically surging and slowing output growth path is positive and approximately linear. The model can therefore support a steady-state growth path. The firm would break-even with the first product line in the sequence, and would make positive and increasing profits as it markets the following ones.

Figure 5.3 Measures of productivity across the product families



In a context where product quality is being improved, productivity growth can be measured in various ways. The left-hand diagram of Figure 5.3 depicts the evolution in quality-adjusted productivity, calculated as the multiple of output and quality divided by labour inputs into

production. The envelope of this variable has a sharp incline close to the initial horizon after which it grows at an approximately constant rate across successive product generations. The diagram on the right depicts the evolution in "revenue productivity," which is measured as revenue deflated by the average price in each product lifecycle (with the price at the very start of the product line normalised to unity). The envelope curve to revenue productivity is steeper than that to quality-adjusted productivity at the outset, but is otherwise very similar. Its path is rapidly increasing and concave in the first product line and about linear over subsequent ones; the latter segment denotes a constant growth rate in revenue productivity across the product families.

Having described the optimised time paths in a scenario where the firm benefits from complete spillovers and market growth, and where it does not discount its profit stream, we now examine how the envelope curves could be shaped in other settings. The results from the comparative dynamics are depicted graphically in the Appendix. Figure A5.1 presents the results from a simulation with no demand-side externalities: the market size is constant. Compared to the benchmark simulation, there is a fall in the level and gradient of the average trends in both classes of R&D; applications-R&D now has a decreasing trend. Naturally, these controls lead to a lower rate of growth in the states of fundamental knowledge and product quality, and the envelope above product quality is concave, not convex. Envelope curves for output, price, revenue, expenditure and profits also have lower levels and less steep inclines. Interestingly, the reduced incentives to fund R&D imply that the break-even is reached sooner, and that the 'fluctuations' in expenditure and hence profits are less marked.

The results obtained for a simulation with positive discounting (δ=.025) in Figure A5.2 exhibit similar declines in the level and slope of the envelope curves. This is to be expected, as this parameter variation would also impact negatively on the anticipated return that can be obtained from a given investment in R&D. Similarly for a specification with

partial spillovers in fundamental knowledge ($\mu_S = .75$). There are, however, a couple of differences worth pointing out in the latter simulation. Figure A5.3 indicates that exploratory R&D expenditures follows a somewhat irregular pattern, being greater in the first cycles but declining in the later ones, as a result of the downward jumps in the state of fundamental knowledge. As the gains in fundamental knowledge in new product lifecycles lag behind over time, the envelope corresponding to the state of knowledge follows a decreasing path. Note that there is no loss in knowledge resources as such: knowledge resources, once obtained, are not misplaced or vanish. The *measured* state of *relevant* fundamental knowledge drops because it is assumed that the knowledge resources are not as appropriate when it comes to the improvement of follow-on product lines.

Other comparative dynamics were run with different parameter values to determine the sensitivity of the numerical results described thus far. The results from these simulations are consistent with those presented in the Appendix. The envelopes curves to quality, output, productivity, and profits are observed to be steeper and more prone to be convex if: knowledge spillovers are more extensive ($\mu_S \to 1$); the market becomes larger with newer product varieties (higher μ_C); the discount rate is lower ($\delta \to 0$); and the production technology is more "flexible" ($\mu_T \to 1$). Naturally, parameters variations that make for steeper evolutions *within* each product line (lower labour costs, more productive research activities, etcetera) would generate steeper envelopes over the M product lines' evolution; see the comparative dynamics analysis in Section 4.4. Conversely, when these parameters are biased in the opposite directions, it is more likely that the extended development process of multiple product lines mimics the individual product lifecycle; i.e., that the envelope to labour, price, output, productivity, and profitability is concave. The general conclusion that can be drawn from the analysis in this section is that knowledge-driven growth processes at the firm-level can exhibit a steady-state or even an increasing path if inter-product-line

knowledge spillovers – of the more fundamental kind pertaining to the research platform or that embodied in the product design – and demand-side externalities are strong enough.

Section 5.4 Optimal number of product lines marketed in a fixed time frame

It is necessary to relax the assumption of a fixed and predetermined time structure to tackle the question of the 'optimality' of the switching decisions. This section supposes that the monopolist wishes to establish the number of product generations that maximise profits when the global time frame is taken as given. Formally, we seek to determine the (integer) number of product lines M^* that maximises cumulative discounted net profits over a fixed horizon $\overline{T}$; that is, the sum of profits minus the fixed costs of switching, all discounted at the appropriate rate. Alternatively, the problem can be conceived as one of selecting the equal time of development for each product family $\tau_i^* = \tau^*$ for all i, subject to the total time being $\overline{T}$.[16] The problem can be framed in this way because, as the comparative dynamics results in Chapter 4 indicated, cumulative profits are monotone increasing in the planning horizon. Using the notation introduced earlier, the problem can be expressed as:

$$\underset{M}{Max} \sum_{i=1}^{M} \left[\int_{(i-1)\overline{T}/M}^{i\overline{T}/M} \delta^{t_i} \Pi\left(N_{Ei}^*(t_i), N_{Ai}^*(t_i), L_i^*(t_i)\right) dt_i \right] - \left[\sum_{i=1}^{M} \left[\delta^{(i-1)\overline{T}/M} F_S\right] - F_S \right]$$

The first term in the expression corresponds to the cumulative discounted profits across the M product lines – before debiting any switching costs. Current profits $\Pi\left(N_{Ei}^*(t_i), N_{Ai}^*(t_i), L_i^*(t_i)\right)$ are computed using the optimal controls in the relevant planning

[16] The firm's choice of planning horizon for its product lines could be motivated by its legal right to a temporary monopoly, such as would be created by a patent on an indispensable research tool or on a component integrated to the different product families, as the end date for its hold on the product market would then be known in advance. Or it may be a result of a firm's effective and long-lasting (but not necessarily legally-grounded) control of key production inputs or distribution channels, in combination with an internal planning process lasting a fixed number of years.

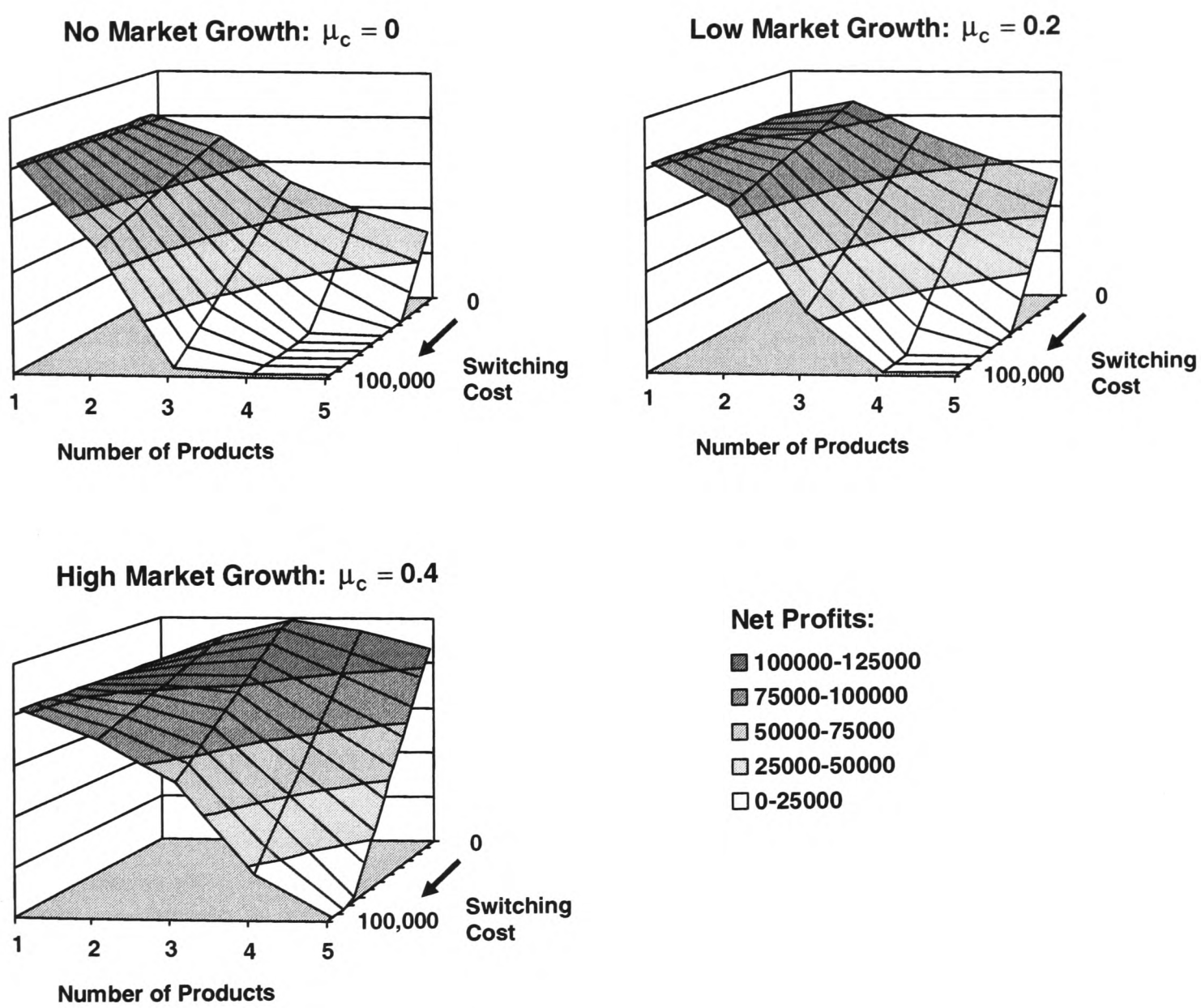
horizons $\{[0,\overline{T}/M],[\overline{T}/M,2\overline{T}/M],...,[(M-1)\overline{T}/M,\overline{T}]\}$. The second term subtracts the discounted fixed costs of the $M-1$ transitions. We solve this optimisation problem numerically by running separate simulations for values of M ranging from 1 to 5, and computing the cumulative profits in each. The total time span is fixed at $\overline{T}=100$. There are complete spillovers of fundamental knowledge $\mu_S=1$, the production technology is equally efficient with any of the product families $\mu_T=1$, and the initial quality of a given product line equals that of the line that is given up $\mu_Z=1$. The discount rate is set at $\delta=.025$. All other parameters are left as in the benchmark simulation of the previous section. The problem is solved for different values of F_S to show the range of equilibria that can be supported in this variant of the model.

The key results of this optimisation are set out in Figure 5.4, which maps out net profits with the number of product lines on one axis, and different values for the switching cost on the other, for three scenarios: one with no market growth; another where market grows by $\mu_C=.2$ when new product lines are introduced; and a final one where the market expansion effect is greater, $\mu_C=.4$. Although this is clearly not the entire matrix of possible outcomes, it is large enough to reveal the nature of the trade-off between extending the time of each product lifecycle against the value that is derived from marketing a greater number of product families.

The top left diagram in Figure 5.4 indicates that (conditional on the other parameter values) profits are strictly decreasing when the market size is fixed, whether or not the transitions are costly. This implies that the firm would select a "corner" solution, in which a single product family is developed and marketed all the way through ($M^*=1$). If the market experiences moderate growth ($\mu_C=.2$), however, *gross* profits (or, alternatively, profits when transitions are costless) exhibit an inverted U-shaped relationship with respect to the

number of product lines M, which peaks at $M=2$. The same is true of profits net of switching costs when these take on low enough values, which implies that the firm would prefer an "interior" solution; the profit-maximising number of products would be $M^* = 2$. If switching costs are higher, net profits become monotone decreasing in the number of product lines for any $M \geq 1$ and a corner solution with a single family would be preferable.

Figure 5.4 Net profits from different numbers of product lines



A comparison of the results with low and high market growth shows that the strength of inter-product-line demand-side externalities plays an important in the level of profits the firm can expect to receive, and the preferred number of product lines. When $\mu_C = .4$, the peak in gross profits is at $M=3$ rather than at $M=2$. As the scale in each diagram is the same, one can

observe that the firm's profits are approximately 20% higher when the market expands at a rate of $\mu_C = .4$. Yet again, low enough values of F_S would lead the firm to market several product families. Other simulations, the results of which are not shown here, indicate that knowledge spillovers have a similar effect to demand-side externalities: the greater the degree of spillovers, the more likely it is that a solution with two or more product families will be optimal.

The modelling approach introduced in this section can be viewed as an extension of the theoretical analysis undertaken by Richard Zeckhauser (1968), "Optimality in a World of Progress and Learning," which also aimed to determine the trade-off between improving one product for an extended period of time vs. switching across a sequence of related products with increasing output potential, but with an intervening cost at each transition. Although the central question asked by Zeckhauser was the same, his framework of analysis assumed that: learning was *exogenous*, depending only on the time during a given product was manufactured; and production was determined by a fixed "rate of flow of labour input." That is to say, Zeckhauser took as given the relationship between the flows of output and revenue, and development time. The only decision variable was the number of products to develop over a fixed horizon, which he derived analytically. In contrast, the relationship between the time of development and the stream of profits is determined endogenously in our numerical analysis, according to the optimised values for the control variables in the dynamic model of heterogeneous investments in R&D.

Both Zeckhauser's model and this one make the restriction of an equal time of development for each of the product lines and assume that the overall planning horizon is exogenous. It is the second of these assumptions that drives the results, as it creates a tension between extending the horizon of each product line, which lets the firm embark on a more ambitious R&D programme and circumvents some switching costs, and marketing several

lines, which in the presence of knowledge spillovers and growing markets may allow it to bootstrap itself onto a steeper growth path. The outcome of this trade-off depends on the strength of the externalities and the pecuniary cost of the transitions, as well as on variables such as the discount rate. In the following section, these restrictions will be relaxed to study accelerated product improvement when the duration of each product lifecycle is variable.

Section 5.5 The financial feasibility of "accelerated product development"

In this section we investigate the R&D and production programme of a firm that aims to shorten the time during which it markets successive product families whilst maintaining its profitability. This seems to be a plausible manner of modelling the problem, as quite a few case studies have observed that one of the objectives commonly required by managers of their development efforts is that they break-even in this sense.[17] As this is the very minimum financial requirement a firm may aspire to, the path derived by the optimisation procedure described below will be called the "financially feasible development path." To be sure, there may be more stringent tests on the product at the gateways in which it is evaluated, such as a minimum level of sales at a given point in time or a particular benchmark for profit margins.[18] Whilst other financial goals that could be set by firms will not be analysed in this

[17] Notwithstanding the publicised enthusiasm of many managers and management consultants, hard statistical evidence about the fortunes of firms shortening their product lifecycles is not easy to find. The problem is one of finding reliable data about the development time and profits in the set of products belonging to the same family, both of which are necessary to test whether the shortening of the cycle actually improves profitability. See Bayus (1998) for a longer discussion of these measurement issues and some evidence on lifecycles and sales (but not on profitability) in the personal computer industry; and also the case study on the portable stereo market by Sanderson and Uzumeri (1995).

[18] On the basis of a survey of managers from Dutch and UK industrial companies, Hart et al (2003, p. 34) conclude that "companies use different criteria at different NPD [New Product Development] evaluation gates." Technical feasibility and market potential seemed to be particularly relevant in the early stages, the focus changing to product performance and project costs during development, and to customer acceptance and sales during and after market launch; those variables in the dimension of market acceptance permeated the entire process. A survey by Tidd and Bodley (2002) of R&D projects in different UK sectors finds that "for both high and low novelty projects the most commonly

chapter, they certainly could be by making minor changes to the optimisation method that is put forward here.

Specifically, the problem is one of finding the minimum time intervals $\{\tau_1^*, \tau_2^*, ...\}$ that allow each product line to generate nonnegative profits net of fixed costs of switching. Denoting the *total* development time for the first j product lines as $T_j = \sum_{k=1}^{j} \tau_k$ for $j = 1,2,...$, one can formally define the minimisation problems as:

$$\underset{\tau_1}{Min} \int_0^{\tau_1} \delta^{t_1} \Pi\left(N_{E1}^*(t_1), N_{A1}^*(t_1), L_1^*(t_1)\right) dt_1 \geq \delta^{\tau_1} F_s, \text{ and}$$

$$\underset{\tau_i}{Min} \int_{T_{i-1}}^{T_{i-1}+\tau_i} \delta^{t_i} \Pi\left(N_{Ei}^*(t_i), N_{Ai}^*(t_i), L_i^*(t_i)\right) dt_i \geq \delta^{T_{i-1}+\tau_i} F_s \text{ for } i = 2,3,...$$

The first equation corresponds to the minimisation problem for the first switching decision, and the problems for subsequent product line transitions are summarily stated in the second expression. In evaluating each successive problem, the financially feasible development time for earlier generations is taken as given. Current profits $\Pi\left(N_{Ei}^*(t_i), N_{Ai}^*(t_i), L_i^*(t_i)\right)$ are derived using the optimal controls in the relevant planning horizons. The simulations assume: complete knowledge spillovers, $\mu_Z = 1$; the other transition parameters equal $\mu_S = \mu_T = 1$ and $\mu_M = .1$; a discount rate $\delta = .025$; and a switching cost $F_S = 5{,}000$.

The procedure by which this sequence of minimisation problems is solved has the following steps. First, we calculate cumulative profits for intervals of time that are multiples of 5 periods, $\{[0,5],[0,10],[0,15],...\}$. After identifying the smallest interval in which profits from the first product line surpass the fixed switching cost, the process is repeated using finer partitions until an integer-valued answer is obtained. Once the time for the first switch has

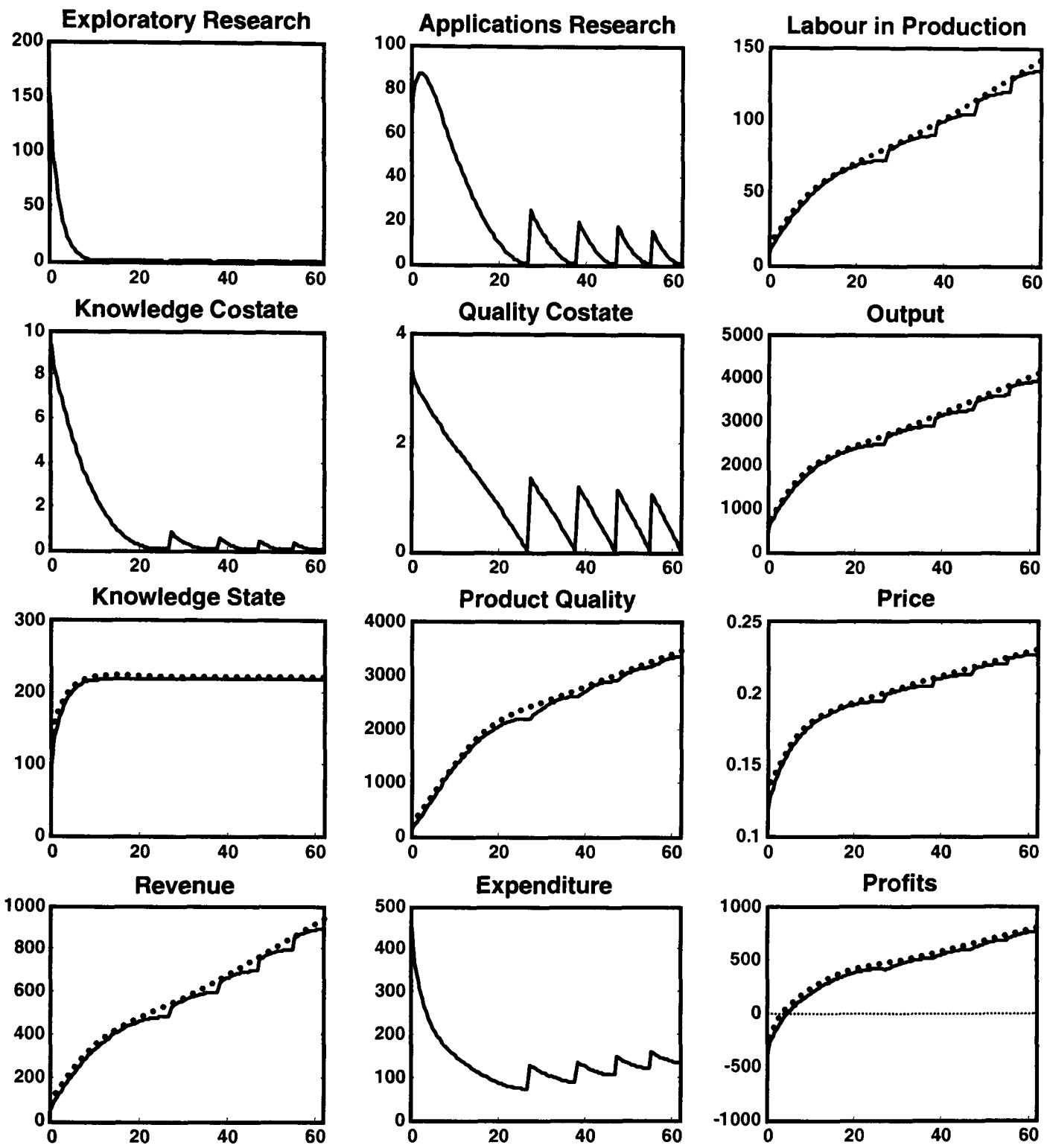
---

used methods for screening and selection were the estimated cost of the project, net present value and payback."

been determined τ_1^*, the minimisation procedure is repeated in a continuation problem, where the initial value of the state variables is derived from the terminal values to the first problem (adjusted by the transition parameters). Initially, the time intervals for the continuation problem are again taken to be multiples of 5 periods $\{[\tau_1^*, \tau_1^* + 5], [\tau_1^*, \tau_1^* + 10], [\tau_1^*, \tau_1^* + 15], ...\}$; the problem is then repeated with smaller intervals. Having obtained the time for the second transition τ_2^*, another continuation problem is set-up going from that time onwards. By computing four continuation problems of this sort, the central properties of the time paths for exploratory- and applications-oriented R&D become clear, as do those corresponding to labour in production and all other relevant variables. The remainder of the section discusses the numerical results from this analysis, presented in a graphical form in Figure 5.5.

The most interesting result relates to the evolving R&D programme, which can be separated into two distinct phases. The first corresponds to the initial product line. This is marketed for a long period, during which the firm makes a considerable investment in both exploratory- and applications-oriented R&D. This is followed by a string of product lines with much shorter development cycles where R&D is predominantly targeted at increasing the quality of the product's design. The envelope curves covering the states of knowledge and product quality, output, price, revenue, and profits are concave. But, in addition, one can observe a sharp rate of change in these variables close to the initial time, when hefty expenditure on both kinds of R&D are generating rapid improvements in product quality. In the phase with shorter-lived follow-on product families, there is an approximately constant upward trend in the envelopes.

Figure 5.5 A financially feasible development path for a product family



The condition of financial feasibility within each cycle of product development operates more forcefully for the first product family because this has a relatively low quality to start off. For this reason, it attracts a weak flow of revenue, and a long programme of development and production is needed to cover the fixed cost of switching to the next family. Much more revenue accrues to the superior design of subsequent product families, allowing the firm to break-even sooner. The results from our framework can help to explain the theoretical basis for one of the 'hidden costs' of "accelerated product development" discussed by Crawford

(1992), in which the firm neglects its search for fundamental information. If the firm were to appreciably shorten its first cycle of product development, the detrimental effects on its product and research platform (working via the reduced incentives to fund exploratory research with short-lived planning horizons) would have long-lasting effects on the development of later product lines, reducing the firm's potential productivity growth and setting back its financial break-even point. Other potential costs of accelerated development, such as the added market and technological risk faced by a less-informed firm or the inefficiencies provoked by the extra pressure on research teams, suppliers, or customers, are not captured by the present modelling approach.

Section 5.6 Conclusions

This chapter has been concerned with the analysis of a profit-maximising firm's programme of heterogeneous research and production across multiple "product lines" or "product families" sold in a product market sheltered from competition. The novelty of the present approach lies in its analysis of R&D spending and production decisions in a context where there are inter-product-line knowledge spillovers and demand-side externalities. Up to now, the related literature had provided many conceptual schemes within which to think about the activities underpinning product family development; see, e.g., Uzumeri and Sanderson (1995) and Halman, Hofer and van Vuuren (2003). But rigorous theoretical analyses with solid microeconomic foundations are far and in between. The contributions by Mahajan and Muller (1996) and Zeckhauser (1968), briefly discussed in the introduction and at the end of Section 5.4 respectively, finesse the firm's investment decisions by assuming that there is an immediate and exogenously given rise in the payoff stream from introducing a newer product generation. What this chapter has attempted to provide is a solution to the firm's product line

management problem that is explicitly built on a dynamic model with endogenous exploratory- and applications-oriented R&D investments.

For the sake of simplicity, we first assumed that each product line was marketed for the same, exogenously given time. Simulations employing a continuation method indicated that the envelope curves framing the progression in quality, output, productivity and profitability had a positive slope and could be convex – in other words, these variables had positive rates of growth and under some parameter values this rate increased over time. An accelerating time path was more likely with low discount rates, full spillovers of fundamental knowledge across successive product generations, flexible production systems, or when newer product lines elicited greater expansions in the consumer base.

The assumption of a predetermined time structure was relaxed later in the analysis to examine the benefits to the firm of making more or less product lines commercially available within a fixed interval of time. The nature of the trade-off (in profit terms) between marketing a small number of longer-lasting product lines or many in rapid succession was seen to depend on the fixed cost of switching between them and other structural conditions. High switching costs could lead to a "corner" solution, in which the firm only markets one product line; with lower costs, a profit-maximising solution with several product lines was feasible. As a last exercise, this chapter assessed the minimum period during which successive product lines needed to be marketed if they were to break-even; i.e., allowing each to finance the cost of switching from its own profits. The financially feasibly time path featured a much longer period of development for the first product family, the only with a substantial investment in exploratory R&D, and frequent replacement of later families, helping to expose some of the limits to "accelerated product development."

This modelling effort was originally motivated by the tension between documented firm-level experiences showing stable or increasing research spending and the theoretical cap

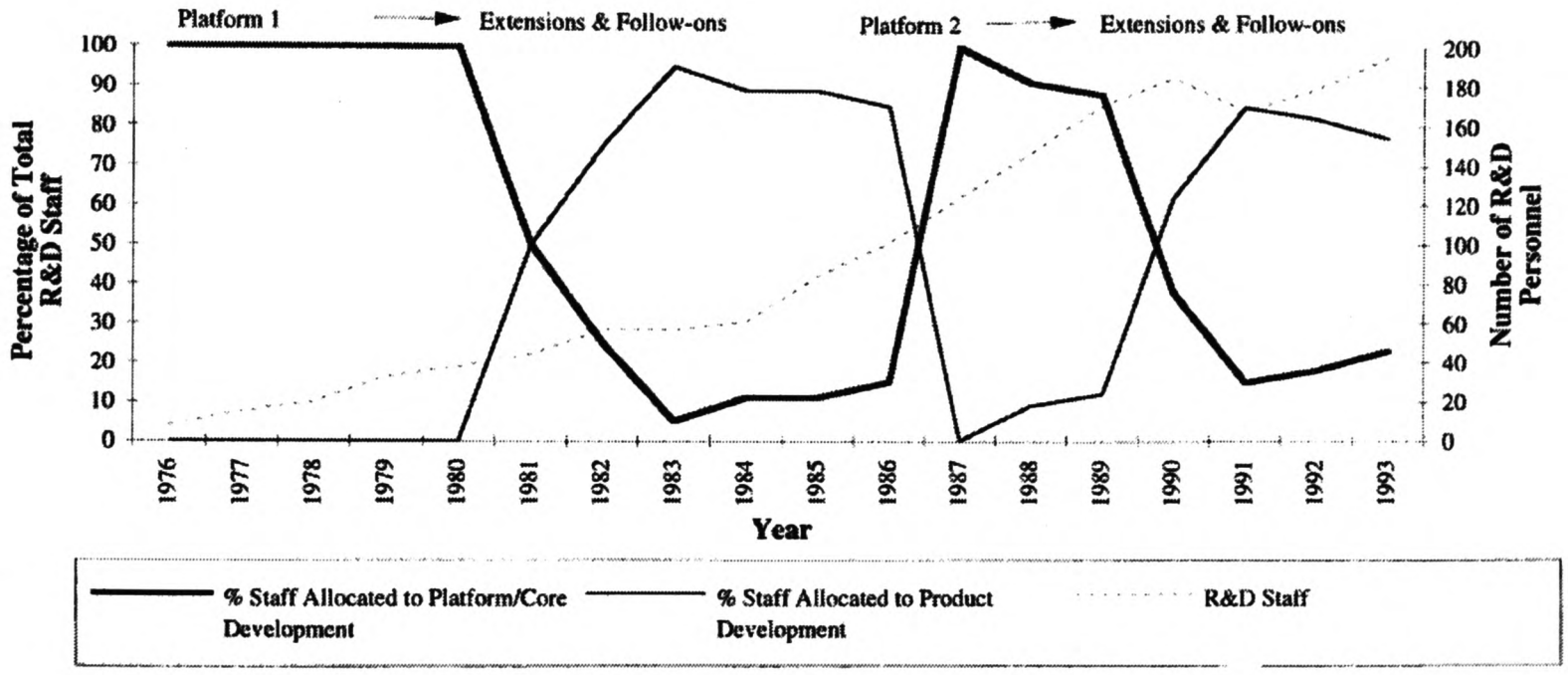
to R&D funding that was encountered as the profit-maximising solution to a dynamic model with a single cycle product development. Chapter 3 suggested that the tension was just as acute if not more so if that model was reinterpreted in aggregate terms, as the outcome of the innovation dynamics at the level of an industry or an entire economy. The approach undertaken here reconciles theory and evidence by focusing on the increasing returns that may be set off by knowledge spillovers and/or external market growth across multiple product lines. Like the AK model or other endogenous growth models, our theoretical set-up can generate permanent output and productivity growth. In contrast to the predictions of new growth theories, however, these variables would not grow in a smooth steady-state fashion over a longer time interval. Instead, sustained growth rests on the evolution of interrelated technologies (or industries) that displace each other over time: growth takes the form of a traverse across individual growth paths, each with limited possibilities of improvement. It is the "envelope" that can be drawn above the individual trajectories that can maintain constant or accelerating growth rates in output or productivity

Numerous examples documented by the technological forecasting literature agree with the results outlined. Figure 5.1 depicted the evolution in the maximum speed attained by different transport modes, a case study due to Robert Ayres (1966). Similar progressions were traced by Ayres (1968) concerning the rate of increase of operating energy in particle accelerators, memory capacity in early computers, and efficiency of external combustion energy conversion systems; other technologies were examined in Ayres (1969). More recent evidence on the development trajectories in successive product life-cycles could be accounted for by the model we have introduced; for example, Sahal (1985) concerning propeller and jet aircraft or Grübler (1990) on various transport modalities.

One can find correlates of this process at other levels of aggregation. There are case studies which show that firms developing successive product families aimed at the same

market see their output and sales go through a sequence of escalating S-shaped paths; e.g., a study on different generations of Canon printers by Matsumoto et al. (2002). The results concerning R&D expenditures on "product platforms" and on the families of products stemming from these platforms are empirically consistent with firm-level time-series data examined by Meyer et al (1997). As an illustration, Figure 5.6 reproduces a figure from their work that charts R&D allocation towards platform and product development for a firm producing "highly complex imaging products used for measurement applications." It would be interesting to collect and analyse more data of this type to investigate the R&D spending patterns in industries with different cost and demand conditions.

Figure 5.6 Platform versus follow-on allocation of R&D staff for a product family[19]



The dynamic model introduced in this chapter and the methods by which it was solved could be elaborated and improved in many respects, of which we briefly consider three in the last part of this section. To start with, the analysis (in this chapter and in the two preceding ones) assumed that the labour supply of workers skilled in production or research was perfectly elastic. In a model seeking to explain the optimisation of one innovating firm over a relatively

[19] Source: Figure 9 in Meyer et al (1997, p. 105).

short horizon, this may be a reasonable approximation. It is harder to justify if a firm's decisions span a very long period of time, as labour market conditions are liable to change. Nor would one wish to make this assumption when construing the results as the outcome of an aggregate R&D process – even if this simplification is accepted for some industries or sectors, it would not be valid generally, and it would be untenable for a country taken as a whole. At these levels of aggregation, numerous studies have found that the labour supply is inelastic, and that of research workers is more so. This reflects the greater scarcity of specific skills required for research activities and the time taken to acquire them.[20] The simplest way to model this labour market feature is by introducing an inclined wage curve, in which further demand for employees drives up the equilibrium wage. The additional labour costs are likely to reduce R&D spending; and one may expect a decrease in output growth rates as well, as a direct consequence of the increase in the wage bill of labour in production, and because of the less steep rate of incline in quality.

Secondly, supply- and demand-side externalities have been premised to be constant across any pair of product families. In reality, one would expect the externalities across new

[20] Over forty years ago, motivated by the experience of "dynamic shortages" of R&D workers they observed in the US during the 1950's, Arrow and Capron (1959, p. 303) produced a study where they suggested that:

> "The elasticity of the supply of engineer-scientists with respect to price changes may be expected to be small but not zero over short periods of time, owing to the length of time it takes to train new personnel. Over longer periods, higher salaries will certainly elicit a greater supply, though again because of the importance of non-economic factors in choosing a career and because of the uncertainty of rewards in the distant future, the responsiveness of supply will be less than for commodities such as manufactured goods."

Even if the theoretical case for low elasticities in the supply of research workers is almost self-evident, actually measuring elasticities to policy shocks such as increased government expenditures is plagued with econometric pitfalls. There exist problems of identification, as supply and demand functions are estimated from observed data that reflect the equilibrium outcomes of adjustment processes between supply and demand, nonstationarity of macroeconomic variables because of time lags, and measurement errors (in the number of working hours of each employee, for example). It was necessary to wait for a revolution in econometric methods and increasing data quality to provide 'hard evidence' on the matter. A careful study by Marey and Borgans (2000) exemplifies the advances made in this field. They employ dynamic econometrics to estimate the wage elasticity of the supply of R&D workers from Dutch macro time series, finding this to be in the region of 0.96 in the short run and 1.20 in the long run. In a small economy like the Netherlands, the elasticity of supply is likely to be quite high, and this estimate is higher than Goolesby (1998) finds for the U.S..

families to be variable and uncertain, and that this random element would form an integral part of the decision to switch or not. This issue could be addressed by introducing a stochastic process for the characteristics of new product families – to what extent the available knowledge base would spillover, the loss in productive efficiency that ensues, and by how much the market would be enlarged. One way of framing the problem in that setting is as follows. In every period, a realisation is drawn from distributions for these characteristics and used to compute the expected sum of discounted profits if the firm switched to the new family. By comparing it to the value from marketing the current family, one could evaluate whether the firm should replace its product family or wait for another time period and a new parameter set.[21] This process could be continued for a predetermined amount of time or until the firm was observed not to switch for a specified number of periods. The model would need to be iterated to estimate the properties of the distributions for relevant decision parameters. Some properties one would particularly want to extract from this simulation are the mean product cycle duration, the average number of product families introduced, and the trend in R&D expenditures across all the families developed.[22]

Lastly, it would be interesting to model the "fully optimal" control problem, in which the firm anticipates the knowledge spillovers and demand-side changes across product lines. In this chapter, the time paths were optimal within each product lifecycle, and the connection between product lines was via the state variables and parameters. Apart from these links, each cycle is assumed to be independent of the next. Consequently, the firm's controls do not account for the full future value of R&D on research and production capabilities. The

[21] An almost identical process is described by Joan Robinson (1962) in her *Essays in the Theory of Economic Growth*, where the firms observe the characteristics of the innovations, in that case relating to the capital-labour ratio and their relative productivity, and decide whether it is worthwhile to adopt the newest one or not.

[22] The alternative is to employ stochastic optimal control to obtain some general properties about the optimal time paths. The basic set-up would need to be adapted to apply the results in the literatures on investment under uncertainty à la Dixit and Pindyck, stochastic control à la Kendrick, or recent contributions on 'multi-armed bandit' problems. The three approaches are listed in terms of increasing mathematical sophistication; see Brezzi and Lai (2002) for a clear exposition of the most difficult one.

solution of the problem in which the firm fully *internalises* the effects of its R&D investments in successive product lines presents a challenge, as it requires a nontrivial extension of the maximum principle.[23] This class of dynamic problems is not at all easy to evaluate formally, and at present there do not seem to be any ready-made research tools to make a numerical analysis. Given the present 'state of the art' in numerical simulation methods, it would be more practical to keep the framework set out here and introduce an additional component in the firm's objective functional giving the future value of the array of knowledge assets gained during development. This "scrap function," as it is called in optimal control theory, would need to closely correspond to the anticipated gain in profits if the model is to remain mathematically consistent. An appropriate functional form could be determined from the comparative dynamic results in Chapter 4. One would then need to derive the new transversality conditions and solve the resulting dynamic system with the numerical continuation method presented in this chapter.

This completes the formal and numerical analysis of the dynamic model of investment in exploratory and applications-oriented R&D by a firm that has a monopoly position in the market for its products, and hence for the innovative products and production methods that such investments will yield. In the next three chapters, we turn to consider a game theoretic analysis of the evolution of oligopolistic industry structure that is driven by the consequences of research investments made under conditions of fundamental uncertainty.

[23] Leonard and Van Long (1992, pp. 310-32) outline the necessary and sufficient conditions for a problem with jumps in the state variables, conditions originally supplied by Vind (1967).

Appendix

Figure A5.1 Development of a product family with constant market size

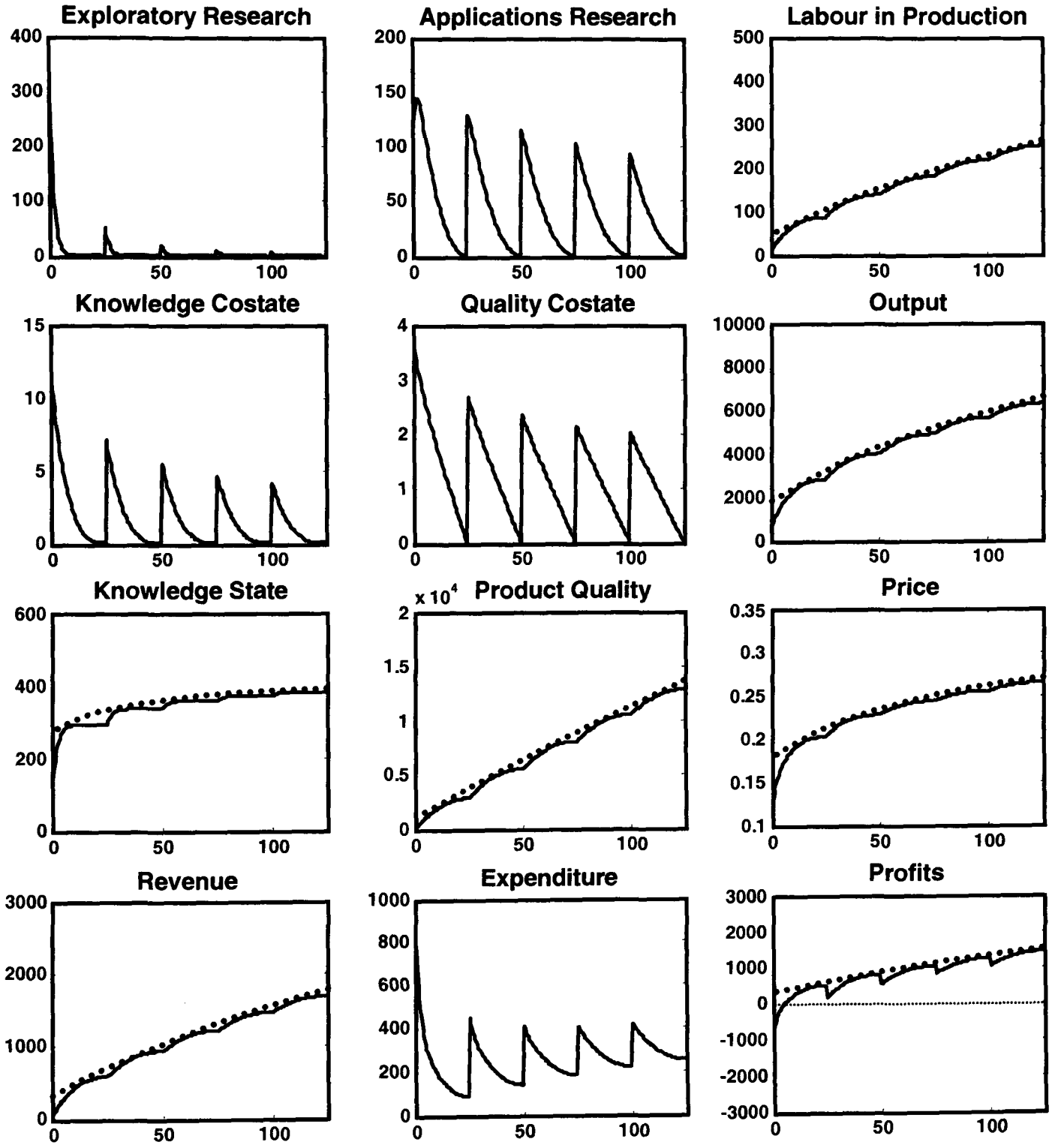


Figure A5.2 Development of a product family under discounting

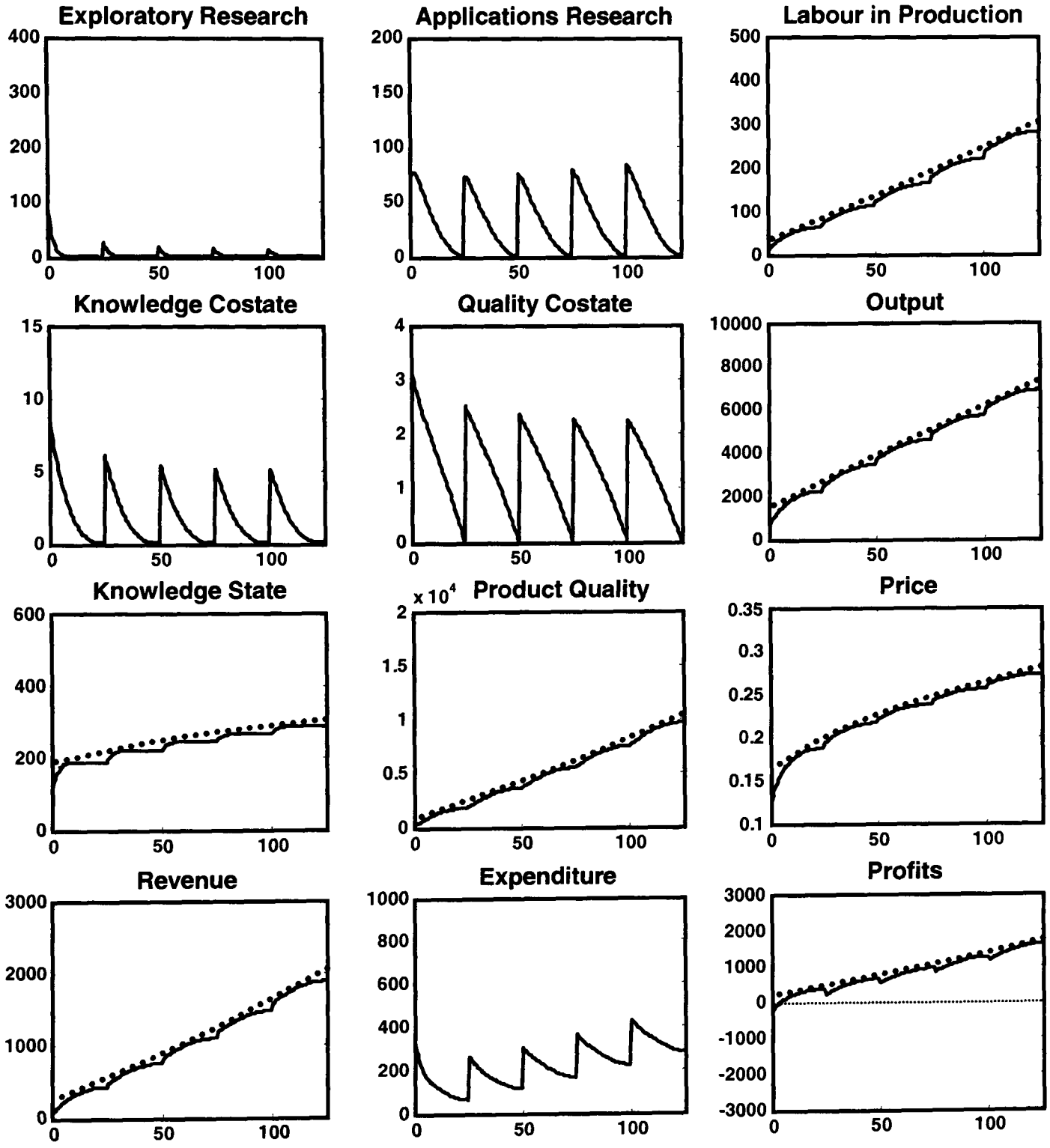
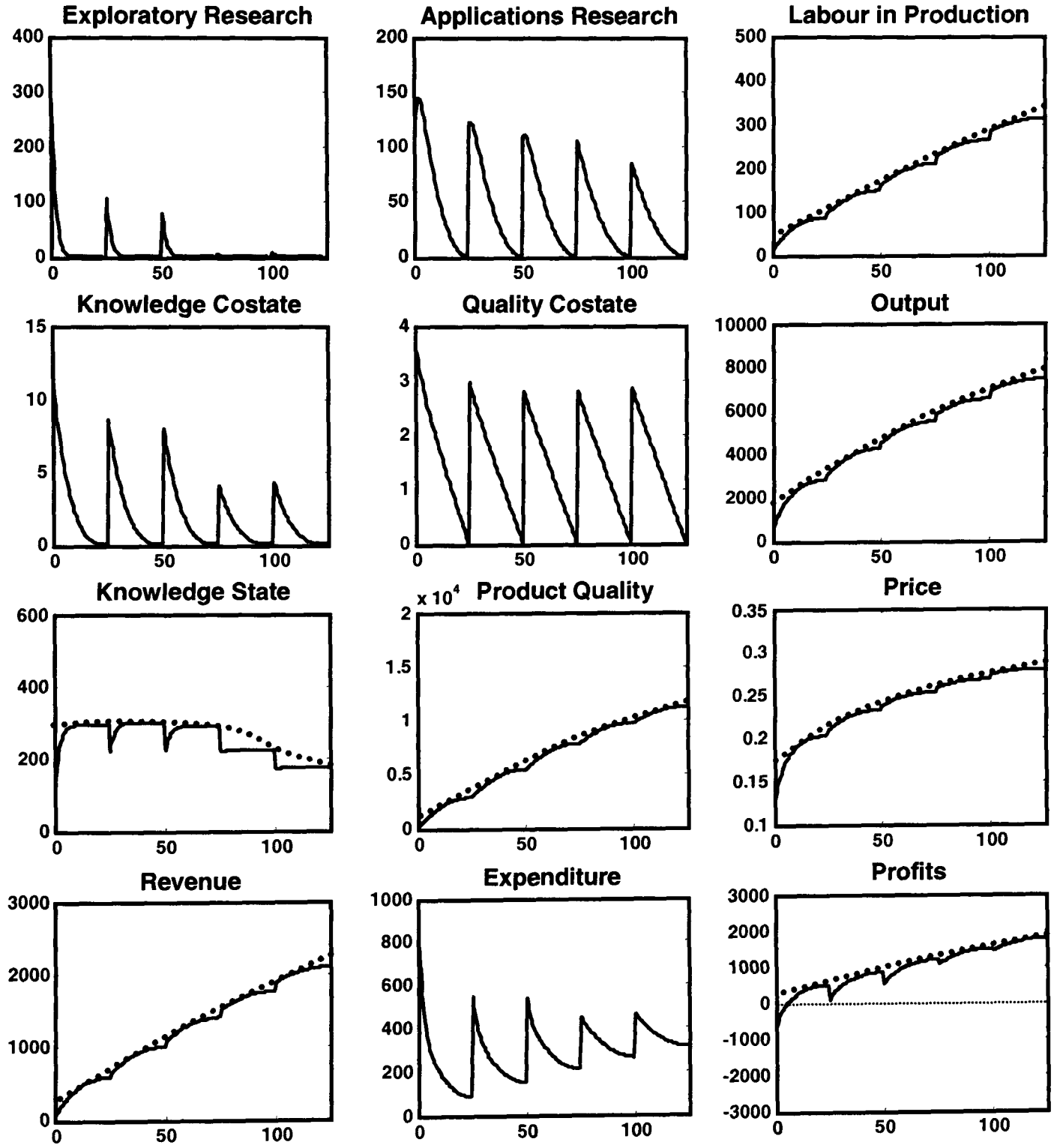


Figure A5.3 Development of a product family with partial knowledge spillovers



Chapter 6

Uncertainty, innovation, imitation and the evolution of market structure

6.1 Introduction

Chapters 3 to 5 investigated the optimal dynamic structure of a firm's exploratory- and applications-oriented research outlays and its associated production decisions when it has succeeded in monopolising a specific product market – as a result of early patenting, or through its control of key production inputs or distribution channels. The ambitions of the enquiry were limited by the extreme position assumed in those chapters: they sought to show how a profit-maximising company would integrate its heterogeneous research activities with its output and pricing decisions, but only when it had a secure hold on the market in question. This allowed the dynamic optimisation model and its extensions to be framed without reference to strategic concerns, the actual or potential threat from substitutes developed and sold by competitors. While extreme, this is not a case without interest, as firms often find themselves in this privileged position, at least temporarily. Some concrete examples are furnished by DuPont's development and sales of neoprene, nylon and Teflon. As Hounshell and Smith (1988, p. 483) point out in their comprehensive history of R&D at this chemical conglomerate, "Teflon was as novel a substance in the 1950s as synthetic rubber and fibers had been in the 1930s. As with these earlier innovations, Du Pont had secured a strong proprietary position, so the company could develop and commercialise Teflon without paying much attention to competitors' activities."

However, the history of R&D at DuPont reveals this company's vulnerability in other product markets, which had to be split up early on with other innovating firms that

independently developed close substitutes, or where there was continued competition from existing products. This occurred in the case of Delrin, another plastic developed by DuPont, about which Hounshell and Smith (1988, p. 491) write: "Delrin was a magnificent technical achievement and eventually became a commercial success, but the early emergence of competition prevented it from becoming a highly profitable product." Moreover, the historical evidence provided by the authors shows that DuPont had different instruments to delay the entry of substitutes, but it could not forestall competitors indefinitely. In the long run, the company's market share usually declined as the product became "commoditised," following the entry of "product clones" produced by competitors that had invested in costly imitative research or had waited for the original patent to expire. A number of empirical studies using different methodologies appear to confirm that in the long-term firms have limited market power in any given product market. In a now classic study of a sample of patented product innovations in the chemical, drug, electronics and machinery industries, Mansfield, Schwartz and Wagner (1981) estimated that sixty percent of the innovations were imitated within four years, long before the patent protection actually expired.[1]

And, of course, not all of DuPont's new product or process research was a success in technical terms. There was a great deal of uncertainty about whether the original targets could be achieved within a reasonable time frame. This did not deter the company from funding this class of research, as it perceived that the return from the "star products" more than covered the losses sustained when their research efforts 'missed the mark.' In a more recent scrutiny of this conglomerate's experience in research, Hounshell (1998, p. 13) writes: "Most certainly, DuPont's executives knew that the returns on short-term, more precisely focused R&D investment were both more certain and more easily measured. But when the company

[1] Naturally, the speed at which technological information leaks out to rival firms is faster than the time taken to begin production of substitutes. In a survey of 100 U.S. firms in major manufacturing industries, Mansfield (1985) found that it generally took between 12 and 18 months for the information to leak out.

moved its research objectives out on the horizon just a little bit, uncertainty grew rapidly. Further out on the horizon, the only certainty was that if the company did not do research, it could not remain competitive." But, needless to say, other players in the chemical industry had no incentive to 'clone' mediocre achievements, for which profit margins were very narrow or even negative.

This chapter turns to consider the interrelated challenges of modelling the irreducible elements of uncertainty in the innovation process and it's bearing upon the dynamics of market competition and industry structure. A novel game theoretic approach is introduced to investigate the evolving industrial structure arising from successive waves of innovation and imitation, in an aim to capture some key aspects of Joseph Schumpeter's 'vision' of the dynamics of technologically-driven development.[2] In the multistage model set out in this chapter, R&D investments in product development are made sequentially, starting with a stage in which firms must decide whether or not to fund new product research, the outcome of which is uncertain in the Knightian sense. The decision process is represented by a subjective state-contingent pay-off matrix, in which the pay-off to alternative strategies depends upon the subsequent evolution of firm behaviours once the fundamental uncertainty is dispelled. Other firms wait until the uncertainty dissipates before investing in costly but certain imitation of the improved product design. These decisions are made in the context of vertical and horizontal product differentiation, in which both innovators and imitators face effective competition from existing (or potential) non-innovating firms manufacturing a conventional product. The specified demand structure is such that firms competing in output with goods of different qualities can coexist in equilibrium making nonnegative profits.

[2] Schumpeter's idea of the "perennial gales of creative destruction," which is described in his famous treatise from 1942, *Capitalism, Socialism and Democracy*, has been reviewed and compared to the Schumpeterian growth theory of Aghion and Howitt (1992) in the second section of Chapter 2.

This is an ideal formulation within which to gauge the value to firms of research strategies motivated by the desire to improve their own products. The latter strategy does not conflict with, but is nevertheless distinct from the use of R&D as a means of deciphering the practical knowledge embodied in other firms' innovations (reverse engineering), or alternatively of reconfiguring the firms' products to circumvent any legal restrictions that would arise from the existence of patents held by others. These are costly learning processes that must be financed by imitators, and, therefore, we may conveniently label them as "imitative-R&D" strategies, to distinguish them from the foregoing "innovative-R&D" strategies. Although they too may yield technical knowledge spillovers, their treatment in the present model departs from the common practice in the literature, in that they are viewed as being neither automatic nor costless, nor immediate in their effects upon the evolving product market equilibrium.

Assessing the value of the two development strategies that will be open to the firms *ex ante* requires a careful analysis on the basis of their expected costs and rewards. Investment decisions have to consider the different sunk costs that need to be financed for new product development and imitation of pre-existing innovations. Whatever the quality of the good they are manufacturing, or the manner in which they obtained the product design, firms face the same fixed cost of production, which they must pay in each stage, and symmetric marginal costs. On the other side of the balance sheet, innovators benefit from additional revenue by competing with a better product before it becomes widely available to competitors, the so-called 'product lead time' effect. This prospective gain is not without its downside, as the fundamental uncertainty surrounding the technological innovation process implies that the magnitude of the 'jump' in quality (and hence revenue) may be vital in assuring an innovator with sufficient supernormal profits to pay for the sunk costs – particularly in the face of unremitting competition from manufacturers of lower-quality products and subsequent entry

by imitators. Seen from this angle, imitation is not only a cheaper form of product development, but also a way of escaping from the uncertainty of innovating.

The explicit examination of "Knightian uncertainty" marks a departure from the mainstream literature studying innovation. To introduce the notion of fundamental uncertainty without abandoning the characterisation of firms as pursuing profit maximisation we study a setting in which there is "imprecise risk."[3] Concretely, suppose that own-product R&D is always expected to result in a product innovation. Firms expect two possible outcomes *ex ante*: a major innovation of quality Z_H or a minor innovation of quality Z_L, with $Z_T < Z_L < Z_H$, where Z_T is the quality of the "standard good" with which all firms are endowed to start with. The state of knowledge only lets firms estimate that the probability of getting a major innovation p lies in the range $\left[\underline{p}, \overline{p}\right]$ where $0 < \underline{p} < \overline{p} < 1$, and that the corresponding probability of a minor innovation $1-p$ lies in the range $\left[1-\overline{p},\ \ 1-\underline{p}\right]$. In other words, the firms' information (of the researchers capabilities, the technological opportunities, etcetera) is insufficient to form an accurate empirical risk assessment, but they can estimate that the measurable risk lies within certain bounds. The intervals $\left[\underline{p}, \overline{p}\right]$ and $\left[1-\overline{p},\ \ 1-\underline{p}\right]$ can be interpreted as a measure of the irreducible Knightian uncertainty surrounding this particular product development project; the further apart the lower and upper bounds are, the more difficult it is to pinpoint the 'true' likelihood of a project's outcome (supposing it existed).

Now suppose that firms pick point-estimates p^* and $1-p^*$ within these intervals to make a calculation of the expected profits from innovative-R&D. This is a subjective

[3] This is a decision-theoretic approach in which agents do not precisely know the probabilities of different states. It has also been called "ambiguity," to separate it from the more radical notion of fundamental uncertainty in which agents are completely ignorant and cannot make even a vague estimate of the relevant probabilities. See Section 2.3 of Chapter 2 for a discussion of the relationship between these concepts and Knight's ideas on risk and uncertainty.

decision that, to fix ideas, can be thought of as based on the optimism or pessimism of the managers and researchers about the probable outcomes – naturally, a different set of underlying factors may be driving their degree of confidence in any specific project they are evaluating. To the extent that the firm is optimistic about the chances of the project's success, it will choose a value nearer the upper bound of the range $\left[\underline{p}, \overline{p}\right]$; when it is dubious of its research efforts, or more prudent, the estimate will be proximate to the lower bound. For the sake of parsimony, the model adopts the following, symmetric belief structure: all the firms believe $p^* = \overline{p}$ and $1 - p^* = 1 - \overline{p}$ when they are optimistic; alternatively, they all believe that $p^* = \underline{p}$ and $1 - p^* = 1 - \underline{p}$ when they are pessimistic. Through these assumptions, it will be possible to analyse the dynamics corresponding to states of 'systemic optimism' and 'systemic pessimism.'[4]

In addition to symmetry of beliefs, symmetry of innovative-R&D outcomes also will be assumed, each innovating firm ending with a product of quality $Z_R > Z_T$. The realised outcome Z_R may coincide with one of the two states that are employed to calculate their expected profitability, Z_L and Z_H. However, given the Knightian notion of uncertainty adopted here, the outcome Z_R would typically be entirely different to what firms had forecasted *ex ante*. As was noted above, investments in imitative-R&D will yield an exact duplicate of the innovators' product design, and therefore their products also have quality Z_R. Taken together, symmetry of beliefs and R&D outcomes are rather strong assumptions,

[4] With the assumption that all firms are either systematically optimistic or pessimistic, the treatment of randomness in the model could be viewed as being analogous to a risk model. As the concluding section argues, the additional structure serves to shape the game structure and equilibrium in terms of outcomes rather than expected returns, and to take into consideration the divergences between subjective expectations and realised outcomes. It is conceptually more attractive because it can be used to delve deeper into questions about belief formation in evolving industries. This methodology is similar to the one implemented by Vaithianathan, Rigotti and Ryan (2001) and Rigotti (2003), who also analyse the "agents' evaluations of innovative opportunities" in the context of imprecise risk by considering the polar cases of maximal pessimism and optimism.

but useful as a first approximation to this problem. The concluding section discusses how relaxing these assumptions could influence the nature of the solution procedure and results.

In finding the equilibrium investment and entry decisions, this chapter espouses Dasgupta and Stiglitz's (1980a) idea of deriving market structure simultaneously to the development decisions of the firms, under the consideration that the long-run "pace of innovative activity and market structure must be traced to more basic ingredients, such as demand conditions, R&D technology, the nature of the capital market, etcetera."[5] This was in sharp contrast to an earlier literature studying market structure and innovation, which could be grouped around two camps, one claiming that innovation was encouraged by a more concentrated industrial structure, while the other proclaimed that the relevant direction of causality was from the innovation process to the market structure that ultimately prevailed.[6] Regrettably, subsequent studies have been more interested in the empirical verification of Dasgupta and Stiglitz's specified model, which was never supposed to be an end product but the starting point of a different approach to the theoretical analysis of the innovation process, one which brought to the fore the sunk costs entailed by innovation and the simultaneous determination of market structure to firms' R&D decisions. Theoretical work following this original contribution has recognised the importance of sunk costs to firms' incentives to innovate, but most of the models in the innovation literature continue to assume a certain market structure, usually a monopoly or a duopoly, and then examine the relative incentives of firms to invest in the development of new products or processes.

[5] Dasgupta and Stiglitz (1980a), p. 267.

[6] Supporting the first claim, it was argued that: concentration, by limiting knowledge spillovers, allowed for a more complete appropriation of the returns from innovation; since imperfect capital markets would not finance investments in R&D with risky or uncertain payoffs, firms with market power would be better placed to finance these investments by relying on internal resources derived from past profits. The opposite direction of causality would follow because of the influence of new products and processes on the pattern of firm growth in an industry – for instance, in promoting a shakeout of backward firms. See the surveys by Kamien and Schwartz (1975, 1982) for an overview of and references to this literature.

This chapter intends to show that Dasgupta and Stiglitz's proposal has a far wider applicability. By extending the endogenous determination of market structure and R&D investments to a dynamic and uncertain setting, one obtains a powerful new solution method with which to study the relationship between technological change and industrial dynamics.[7] To analyse the problem in this manner, it proves necessary to introduce an equilibrium concept that determines how many firms the market can support in each stage of the evolution, conditional on the revealed outcome of the innovative-R&D investments undertaken. The "symmetric sustainable equilibrium" is defined as the maximum number of symmetric innovating firms that can anticipate making zero profits given their subjective expectations about the uncertain innovation process, and the prospect that traditional and imitating firms will enter up to the point where they get zero profits in each stage; and the maximum number of symmetric imitating and traditional firms that can actually get zero profits at each stage, conditional on prior, equilibrium investments in innovative-R&D and the realised outcome of these investments. The procedure by which the equilibrium will be found combines backward recursion with zero profit and zero expected profit conditions for each type of firm. This solution process needs to take account of the (i) leader-follower structure of the sequential investments in R&D, (ii) entry decisions by traditional producers in each stage, (iii) simultaneous quantity competition from which per-stage payoffs are derived, and (iv) technological uncertainty.

The remainder of this chapter is organised in the following manner. Section 6.2 sets out a specified model, explaining in more detail the timing of the sequential entry game, the treatment of fundamental uncertainty, the cost and demand structure faced by the firms, and

[7] It will be seen that this extension carries the implication that the sequence of entry decisions will matter in determining the equilibrium, unlike Dasgupta and Stiglitz (1980a), where there was but one "wave of R&D investments," and their outcome was certain. The spirit of this analysis also bears some resemblance to the "bounds" approach developed by Sutton (1998), in studying the structure of markets when fixed investment in R&D is an instrument of rivalry among firms.

the equilibrium concept that is necessary in solving the model. It then formalises a solution procedure by which the equilibrium can be found. Section 6.3 undertakes a formal analysis of the model, which shows that: (i) if a symmetric sustainable equilibrium with positive investment in innovation exists, it will be unique; (ii) a wide range of possible "industry evolutions" can be supported, including Schumpeterian surges of "innovative entry" and ensuing episodes of "creative destruction," in which traditional producers are driven out of the industry and innovators' rents are eventually eroded; (iii) there are exact conditions under which certain other "historical evolutions" can be shown to be infeasible; (iv) in limiting cases, small variations in the model's parameters may produce discrete changes in the equilibrium composition of firms at each stage of the evolution. The concluding section reflects on the scope for relaxing some of the assumptions in the model, and explains the reasons warranting a numerical investigation of the equilibrium dynamic sequences of the game theoretic model, the aim of the following chapter.

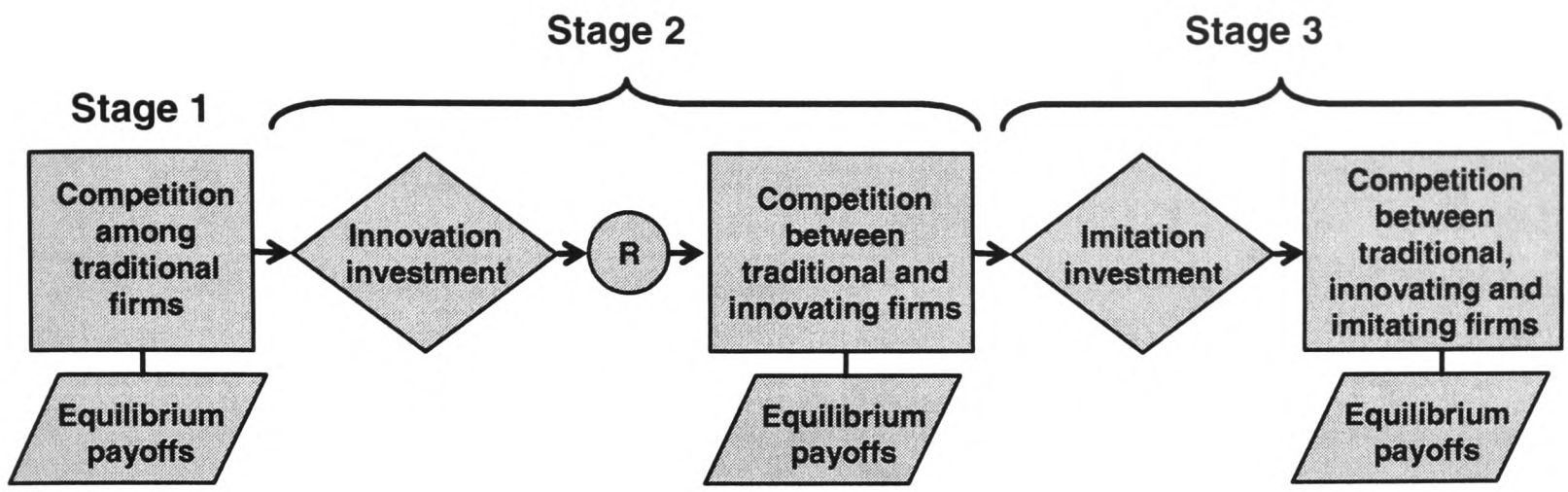
6.2 Industrial dynamics under Knightian uncertainty: A formal model

6.2.1 Notation and assumptions

The evolution of the industry is separated into three stages, as shown in Figure 6.1. The first corresponds to a mature industry, in which traditional producers compete amongst each other by marketing a standard product. In the second stage, firms decide whether or not to invest in fundamentally uncertain product innovation. After the investments are made, the innovating firms find out the outcome of the innovation process and market their new product in a marketplace where traditional producers may be active; the circled "R" in the diagram stands for the stochastic element of own-product R&D. In the final stage, non-innovators can undertake costly but certain imitation activities to learn the improved product design, and

there is a final round of product market competition, this time between the original innovators, entering imitators and traditional producers.

Figure 6.1 The sequential structure of the model



More precisely, the three stages comprise of the following:

Stage 1

- $N_{T,1}$ "traditional producers" compete in quantity with a product of quality $Z_T > 0$. There is free entry of firms wishing to market the "standard product." Beyond the assumption that the industry starts with a 'level playing field,' with costless access to an identical product design, this stage has no other bearing on the calculations of profitability by prospective innovators and imitators. Nonetheless, it is useful as a point of reference against which to compare the evolution of the industry.

Stage 2

- An unlimited number of firms decide whether or not to disburse a fixed sum $F_I > 0$ towards research seeking to develop an improved product. The universe of firms that can make this investment is not confined to the incumbents manufacturing the standard

product: there is a perfect capital market, allowing any number of entrants to readily take advantage of the envisaged investment opportunities.

- The outcome of own-product inventive R&D is fundamentally uncertain. The outcomes are expected to be either a major innovation of quality Z_H or a minor innovation of quality Z_L, where $Z_T < Z_L < Z_H$.
- The state of knowledge only lets firms estimate that the probability of getting a major innovation p lies in the range $[\underline{p}, \overline{p}]$ where $0 < \underline{p} < \overline{p} < 1$, and that the corresponding probability of a minor innovation $1-p$ lies in the range $[1-\overline{p},\ \ 1-\underline{p}]$.
- There is a symmetric belief structure: when 'optimistic,' all firms believe that the probability of developing a major innovation is $p^* = \overline{p}$ and that of a minor innovation is $1-p^* = 1-\overline{p}$. When 'pessimistic,' they believe the probability of a receiving a major innovation is $p^* = \underline{p}$ and that corresponding to a minor innovation is $1-p^* = 1-\underline{p}$.
- After the N_I firms have sunk the fixed cost F_I, they learn the outcome of their research efforts. Symmetry of R&D outcomes implies that each and every one of the innovating firms paying F_I obtains an innovation of quality $Z_R > Z_T$.
- The N_I innovating firms compete in quantity with their improved product against traditional producers.[8] Firms marketing the standard good can freely enter or exit the final product market; the number of traditional producers is denoted as $N_{T,2}(Z_i)$; *ex ante* the firms forecast $Z_i = Z_H, Z_L$, *ex post* they find out that $Z_i = Z_R$.

[8] Notice that this specification of the model supposes that all the firms investing in innovative-R&D compete in quantity in stages 2 and 3. For completeness, the Appendix to this chapter studies how the model would have to change if innovating firms can exit the industry in either stage, which they would want to do when competition with a product of quality Z_R leads to a loss beyond the sunk cost F_I. This is achieved by the inclusion of "ex post participation constraints" for innovators. The extension vindicates the theoretical results we will present in the main body of the chapter, and generates additional findings that are interesting in their own right.

Stage 3

- Any non-innovators remaining in the market and an unlimited number of potential entrants decide whether or not to imitate the improved good by paying a fixed "cloning" cost $F_C > 0$.
- The payment of F_C guarantees imitating firms an exact copy of the product that was developed by innovators. The number of cloning firms is denoted $N_C(Z_i)$, where again $Z_i = Z_L, Z_H$ *ex ante* and $Z_i = Z_R$ *ex post*.
- There is a final round of market competition, this time between innovating, imitating and traditional firms. The number of traditional firms is $N_{T,3}(Z_i)$. There is free entry by traditional producers at this stage also.

To simplify the solution of the model, it is assumed that production costs are the same for all the firms in the industry. Whatever the quality of the good they are manufacturing, or the manner in which they obtained the product design, each firm faces a fixed cost of production $F_P > 0$ and a constant marginal cost $0 < c < 1$; the marginal cost is restricted to this interval because of the demand specification that is employed to solve the model. The fixed production cost is paid in each stage that a firm competes in the final product market; this implies that an innovator planning to sell its improved product in stages 2 and 3 will need to factor in the fixed cost of production at each of these stages. The firms are assumed to discount their profits at a rate of δ, $0 < \delta \leq 1$.

In addition to specifying the timing of the research and production decisions open to the firms, and the fixed and marginal costs these give rise to, it is necessary to stipulate the demand structure faced by the firms. Because one of the aims of the this study is to characterise the evolution of market structure when goods of different quality can co-exist in

equilibrium, we adopt a functional form for consumer preferences suggested by John Sutton (1997, 1998), and employed more recently by George Symeonidis (2000, 2003). The quality-augmented quadratic utility function they propose takes the following form when there are N firms active in the market, each offering a different product:

$$U = \sum_i \left[q_i - \frac{q_i^2}{Z_i^2} \right] - \sigma \sum_i \sum_{j<i} \frac{q_i}{Z_i} \frac{q_j}{Z_j} + M ,$$

where q_i and Z_i are firm i's output and product quality, $\sigma \in [0,2]$ is an inverse measure of the degree of horizontal product differentiation, and M is the expenditure by consumers on outside goods $M = Y - \sum_i P_i q_i$. As Sutton shows, utility maximisation produces a linear demand system, in which the inverse demand function of each consumer for the good sold by firm i is $P_i = 1 - 2q_i/Z_i^2 - \sigma/Z_i \sum_{j \neq i} q_j/Z_j$. When σ equals zero, there is complete product differentiation: the demand for a given firm's product is independent of the others quality or their output. As a result, the equilibrium price and quantity set by each firm would be independent of the other producers' choices. The other extreme is when σ is equal to 2, a situation in which there is no horizontal product differentiation: consumers regard products of equal quality as perfect substitutes.

As the incentive to undertake each class of R&D rests on its prospective profitability, one needs to derive expressions for the profits and expected profits of each type of firm at each stage. To make the notation less cumbersome, it is useful to begin by defining a "gross profit" function. This is the profit made by a particular firm in the final product market after paying for the marginal cost of production, but before subtracting the relevant fixed costs of production or product development. Let $\pi(Z_j | N_T, N_I, N_C, Z_i)$ denote the "gross profit" made by a firm marketing a product of quality $Z_j = Z_T, Z_L, Z_H, Z_R$, given that it is competing in an

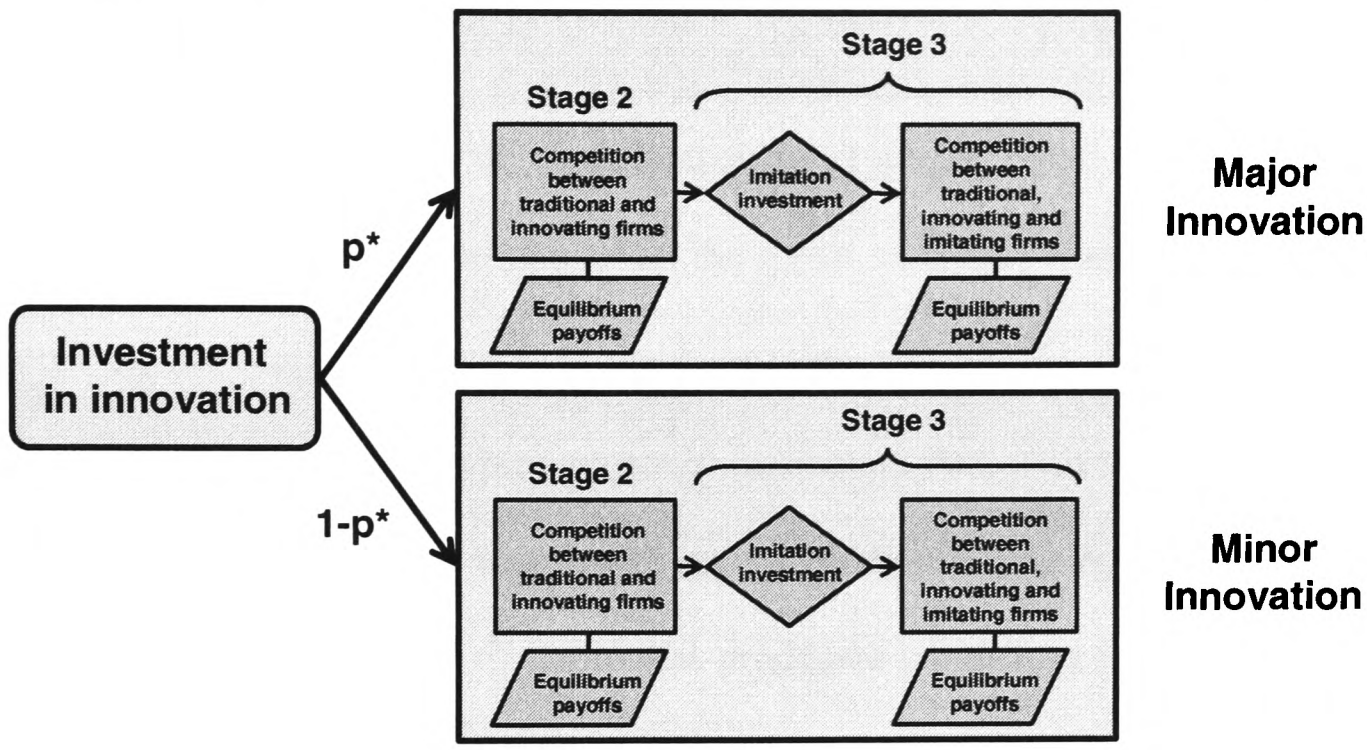
industry with N_T traditional firms, and N_I innovators and N_C imitators producing a good of quality $Z_i = Z_L, Z_H, Z_R$. E.g., the gross profit made by a firm marketing the conventional good in a market with 2 more traditional producers, as well as 1 innovator and 2 imitators that can produce a major innovation $Z_i = Z_H$, would be $\pi(Z_T|3,1,2,Z_H)$; the gross profit of a typical innovator in this case would be $\pi(Z_H|3,1,2,Z_H)$.

By using this function, it is possible to succinctly define the net profits of a traditional producer in stage 1 as $\Pi_{T,1} = \pi(Z_T|N_{T,1},0,0,Z_i) - F_P$, equal to gross profits minus the per-stage fixed cost of production F_P. Net profits in stage 1 will depend solely on the number of traditional producers in the market. In stage 2, a traditional producer earns net profit $\Pi_{T,2}(Z_i) = \pi(Z_T|N_{T,2},N_I,0,Z_i) - F_P$, and at stage 3 this equals $\Pi_{T,3}(Z_i) = \pi(Z_T|N_{T,3},N_I,N_C,Z_i) - F_P$. The second and third stage expressions for net profits are a function of the number of firms investing in innovation, the anticipated or actual outcome of the innovation process, the number of traditional and imitating firms in the market, and the fixed cost of production.

In calculating the expected profits of an innovator, one must first identify the net profits in stages 2 and 3. When the improved product is of quality Z_i, the net profit in stage 2 is $\Pi_{I,2}(Z_i) = \pi(Z_i|N_{T,2},N_I,0,Z_i) - F_P - F_I$, equal to the gross profit minus the per-stage fixed cost of production and the sunk cost of innovation F_I. In stage 3, an innovator obtains $\Pi_{I,3}(Z_i) = \pi(Z_i|N_{T,3},N_I,N_C,Z_i) - F_P$. It therefore would get cumulative, discounted profits equal to $\Pi_I(Z_i) = \Pi_{I,2}(Z_i) + \delta\,\Pi_{I,3}(Z_i)$ when the outcome of innovation has been Z_i. The expected (net, cumulative and discounted) profits of a typical innovator are $\Pi_I = (1-p^*)\Pi_I(Z_L) + p^*\Pi_I(Z_H)$, where $1-p^*$ is the subjective probability that the new

product will be a minor innovation, and p^* is the probability of a major innovation. The value these probabilities take depend on whether they are 'optimistic' or 'pessimistic.' Depicting the evolution of the industry from the point of view of the potential innovators may clarify the logic of these calculations; see Figure 6.2.

Figure 6.2 An innovator's perspective of the industry evolution



An imitator's net profits are $\Pi_C(Z_i) = \pi(Z_i | N_{T,3}, N_I, N_C, Z_i) - F_P - F_C$. These are equal to the gross profits gained in stage 3 by marketing a product of quality Z_i, minus the fixed cost of production and the sunk cost of imitative-R&D F_C. Given the assumption that clones are of equal quality to the original innovation, the difference between the third stage profits of an innovator and the profits of an imitator is that the latter must finance the fixed imitation cost.

6.2.2 The equilibrium concept and a solution procedure

The equilibrium concept with which the model is solved will be called the "symmetric sustainable equilibrium" (denoted hereafter as SSE). It corresponds to the maximum number

of symmetric innovating firms that can anticipate making zero profits given their subjective expectations about the uncertain innovation process, and the prospect that traditional and imitating firms will enter up to the point where they get zero profits in each stage; and the maximum number of symmetric imitating and traditional firms that can get zero profits at each stage, conditional on prior, equilibrium investments in innovative-R&D and the realised outcome of these investments. The equilibrium numbers of firms of each type are denoted as: N_I^*, N_C^*, $N_{T,1}^*$, $N_{T,2}^*$, $N_{T,3}^*$.

The equilibrium number of traditional firms at stages 2 and 3, and the number of imitators need to be evaluated conditional on the actual state they find themselves in, that is, conditional on the realised outcome of the innovation process Z_R. Whereas the nature of uncertainty and the method by which potential innovators deal with this uncertainty govern the equilibrium industry-wide investment in innovative-R&D, aggregate spending on imitative-R&D and the number of traditional producers competing in the market over the industry evolution depend on the relative success of innovating firms' research efforts. In other words, the SSE pins down the patterns of firm entry that would actually come about: N_I^* is determined from firms' forecasts about the innovation process and other firms' future behaviour; the equilibrium number of traditional and imitating firms post-innovation (N_C^*, $N_{T,2}^*$, $N_{T,3}^*$) is calculated retrospectively, on the basis of the equilibrium investment in innovative-R&D and its actual outcome. Dasgupta and Stiglitz (1980a) made use of a related equilibrium concept in solving a model of process innovation, employing a zero profit condition to find the equilibrium in a context of free entry. But to extend their idea to a dynamic and uncertain setting requires some additional thought, as the sequence of entry decisions will matter, and a solution procedure must be established to deal with the 'historical' details of product development.

As the R&D investments occur sequentially in the model, its structure resembles that of a Stackelberg game with innovators as leaders and imitators as followers. From this perspective, traditional producers can be conceived as an extra tier of followers that at each stage in the game enter the market based on the competition expected from firms that previously invested in some variety of product development. The method of backward recursion described below accommodates these two dynamics. There is another feature of the present analysis that separates it from a typical Stackelberg game. Instead of having "reaction functions" for individual firms, the model introduces "sustainable entry" functions that determine the maximum number of firms of a particular type – innovators, imitators, or traditional producers – that can obtain zero profits or zero expected profits. For example, the feasible numbers of traditional producers at the third stage is derived from a zero profit condition that takes as given the number of innovating and imitating firms. This function can then be substituted into the profit function of a typical imitator in the same way as a "reaction function," thereby allowing firms to assess how investments in cloning will influence future market rivalry. This process is repeated for all the different types of firms that follow the lead of the innovating firms, to ultimately get an expression for the expected net profits of a typical innovator as a function of the number of innovating firms N_I. This function is fundamental to the solution because it incorporates all the indirect strategic effects of innovation on imitators and traditional producers, and that derived from the irreducible element of uncertainty in the innovation process.

The solution of the model can be separated into three steps. The first is to get a general expression for the Cournot-Nash symmetric equilibrium profits $\pi^{CN}\left(Z_j | N_T, N_I, N_C, Z_i\right)$. These are the gross profits obtained by a firm manufacturing a product of quality Z_j evaluated at the equilibrium prices and quantities that result from competing in output in a particular product market; i.e., a market containing $N_T \geq 0$

traditional firms, and $N_I \geq 0$ innovating firms and $N_C \geq 0$ imitating firms marketing a good of quality Z_i. The second step in the solution is to substitute the Cournot-Nash gross profit function into the expressions for net profits secured by each type of firm, and then, by using backward recursion in conjunction with zero profit and zero expected profit conditions, to derive the equilibrium number of innovating firms. Lastly, the equilibrium number of innovating firms and the actual quality improvement their researchers have succeeded in obtaining are employed to derive the equilibrium number of traditional and imitating firms in stages 2 and 3. The following subsections formalise these ideas.

Cournot-Nash Equilibrium

The Cournot-Nash equilibrium for a vertically and horizontally differentiated product oligopoly in which consumers have quadratic utility functions was originally derived by Sutton (1997, 1998) and later extended by Symeonidis (2000, 2003). The results reproduced in this section follow the latter author's formulation, but the notation is adapted to fit the special assumptions of the present model. Symeonidis shows that the market demand functions corresponding to the quality-augmented quadratic utility functions yield a closed-form solution for the Cournot-Nash equilibrium outputs and prices in the region where product quality Z_i is positive for all firms $\forall i$, assuming a symmetric fixed marginal cost of production $0 < c < 1$. Let S be the number of consumers, a measure of market size. Then the Cournot-Nash equilibrium output $q^{CN}\left(Z_j | N_T, N_I, N_C, Z_i\right)$ of a firm manufacturing a product of quality Z_j, given that it is competing in an industry with N_T traditional firms, and N_I innovators and N_C imitators producing a good of quality Z_i is equal to:

$$q^{CN}\left(Z_j | N_T, N_I, N_C, Z_i\right) = \frac{S\left[Z_j\right]\left[1-c\right]\left[\left(4+\sigma\left(N_T+N_I+N_C\right)\right)Z_j - \sigma N_T Z_T - \sigma\left(N_I+N_C\right)Z_i\right]}{\left[4-\sigma\right]\left[4+\sigma\left(N_T+N_I+N_C-1\right)\right]}$$

The Cournot-Nash equilibrium price $P^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)$ commanded by a product of quality Z_j in a market composed of N_T traditional firms, and N_I innovators and N_C imitators producing a good of quality Z_i equals:

$$P^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)=c+\frac{2[1-c]\left[(4+\sigma(N_T+N_I+N_C-1))Z_j-\sigma N_T Z_T-\sigma(N_I+N_C)Z_i\right]}{Z_j[4-\sigma][4+\sigma(N_T+N_I+N_C-1)]}$$

The Cournot-Nash price is a mark-up over the marginal cost that depends on the prevailing composition of the industry – i.e., the number of firms in the industry and the quality of their products. Using the equilibrium price and quantity, it is possible to get a closed-form solution for the gross profits of a firm with a product of a given quality. Absent any fixed costs, the Cournot-Nash equilibrium profits $\pi^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)$ will be a function of the expected or realised outcome of the innovation process $Z_i=Z_L,Z_H,Z_R$ and the number of firms of each kind that are present in the industry:

$$\pi^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)=2S\left[\frac{(1-c)\left[[4+\sigma[N_T+N_I+N_C-1]]Z_j-\sigma N_T Z_T-\sigma[N_I+N_C]Z_i\right]}{[4-\sigma][4+\sigma[N_T+N_I+N_C-1]]}\right]^2$$

The expression for firm-level gross profits $\pi^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)$ is increasing in the market size, decreasing in the marginal cost of production, increasing in a firm's own product quality and decreasing in the rival products' quality. The formulas for the Cournot-Nash equilibrium price, output and profits are only valid in markets with two or more firms $N_T+N_I+N_C\geq 2$. If there is only one firm selling a product of quality Z_j, one must derive the equilibrium price and output set by a monopolist, and hence obtain its gross profit function. When consumers have quality-augmented quadratic utility functions, this can be directly obtained from the previous formulas as the special case in which there is complete

product differentiation $\sigma = 0$. Section 6.3 will assume that the parameter values support two or more firms with the relevant qualities in each stage of the evolution in all of its derivations. The following chapter will allow for the existence of a monopoly before any investment in product development is made or its emergence as a result of innovative-R&D.

Backward recursion, zero profit and expected profit conditions

The second step in the solution is to substitute the Cournot-Nash gross profit function $\pi^{CN}\left(Z_j \middle| N_T, N_I, N_C, Z_i\right)$ into the expressions for net profits secured by each type of firm, and then, by using backward recursion and zero profit and expected profit conditions, to calculate the equilibrium number of innovating firms. This part of the solution procedure can be divided into the following sub-steps. For any product quality that could be developed $Z_i > Z_T$, one must:

1. Derive the maximum number of traditional firms that can get zero profits in stage 3, given that there are N_I innovators and N_C cloning firms competing with a good of quality Z_i. The "sustainable entry function" for traditional firms $N_{T,3}^{MAX}\left(N_I, N_C, Z_i\right)$ is derived from the zero-profit condition for this type of firm:

$$\Pi_{T,3}(Z_i) = \pi^{CN}\left(Z_T \middle| N_{T,3}, N_I, N_C, Z_i\right) - F_P = 0$$

2. Substitute the sustainable entry function $N_{T,3}^{MAX}\left(N_I, N_C, Z_i\right)$ into the net profits of a typical imitator, and set this equal to zero:

$$\Pi_C(Z_i) = \pi^{CN}\left(Z_i \middle| N_{T,3}^{MAX}\left(N_I, N_C, Z_i\right), N_I, N_C, Z_i\right) - F_P - F_C = 0$$

The maximum number of imitators that can get zero profits once each has factored in the anticipated competition from traditional producers is $N_C^{MAX}(N_I, Z_i)$.

3. Similarly, derive the maximum number of viable traditional producers in the stage before product clones have entered the market $N_{T,2}^{MAX}(N_I, Z_i)$ from the profit functions of traditional producers in stage 2. The relevant zero profit condition is:

$$\Pi_{T,2}(Z_i) = \pi^{CN}\left(Z_T \middle| N_{T,2}, N_I, 0, Z_i\right) - F_P = 0$$

4. $N_{T,3}^{MAX}(N_I, N_C, Z_i)$, $N_C^{MAX}(N_I, Z_i)$ and $N_{T,2}^{MAX}(N_I, Z_i)$ are substituted consecutively into a typical innovator's payoff until this depends solely on the number of firms that are innovating N_I and the state of nature Z_i:

$$\Pi_I(Z_i) = \Pi_{I,2}(Z_i) + \delta\,\Pi_{I,3}(Z_i)$$

$$= \pi^{CN}\left(Z_i \middle| N_{T,2}^{MAX}(N_I, Z_i), N_I, 0, Z_i\right) - F_P - F_I$$

$$+ \delta\pi^{CN}\left(Z_i \middle| N_{T,3}^{MAX}\left(N_I, N_C^{MAX}(N_I, Z_i), Z_i\right), N_I, N_C^{MAX}(N_I, Z_i), Z_i\right) - \delta F_P$$

The expected profits of a given innovator equal the weighted sum of the discounted cumulative profits anticipated in each of the two states of nature, where the weights are the firms' subjective probabilities about the innovation process. The maximum number of innovating firms that can expect to make zero profits N_I^* can be derived from the following equation:

$$\Pi_I = (1 - p^*)\Pi_I(Z_L) + p^*\Pi_I(Z_H)$$

$$= (1 - p^*)\left\{\pi^{CN}\left(Z_L \middle| N_{T,2}^{MAX}(N_I, Z_L), N_I, 0, Z_L\right) - F_P - F_I\right.$$

$$\left. + \delta\pi^{CN}\left(Z_L \middle| N_{T,3}^{MAX}\left(N_I, N_C^{MAX}(N_I, Z_L), Z_L\right), N_I, N_C^{MAX}(N_I, Z_L), Z_L\right) - \delta F_P\right\}$$

$$+ p^{*}\left\langle \pi^{CN}\left(Z_H \middle| N_{T,2}^{MAX}\left(N_I, Z_H\right), N_I, 0, Z_H\right) - F_P - F_I \right.$$

$$\left. + \delta\pi\left(Z_H \middle| N_{T,3}^{MAX}\left(N_I, N_C^{MAX}\left(N_I, Z_H\right), Z_H\right), N_I, N_C^{MAX}\left(N_I, Z_H\right), Z_H\right) - \delta F_P \right\rangle = 0$$

Unlike $N_C^{MAX}(N_I, Z_i)$, $N_{T,2}^{MAX}(N_I, Z_i)$ or $N_{T,3}^{MAX}(N_I, N_C, Z_i)$, N_I^* is not a "sustainable entry function." N_I^* is the number of innovators in a SSE, which depends on the value taken by the probabilities, and all other specified parameters in the model.

Completing the SSE

By substituting N_I^* and Z_R into $N_C^{MAX}(N_I, Z_i)$ one obtains the equilibrium number of imitators that are supported in the market $N_C^* = N_C^{MAX}\left(N_I^*, Z_R\right)$. The equilibrium number traditional producers in stage 3 is similarly calculated by substituting N_I^*, Z_R and N_C^* into $N_{T,3}^{MAX}(N_I, N_C, Z_i)$, to get $N_{T,3}^* = N_{T,3}^{MAX}\left(N_I^*, N_C^*, Z_R\right)$. The equilibrium number of traditional producers in stage 2 is established by substituting N_I^* and Z_R in $N_{T,2}^{MAX}(N_I, Z_i)$, $N_{T,2}^* = N_{T,2}^{MAX}\left(N_I^*, Z_R\right)$. The pre-innovation equilibrium in stage 1 plays no part in the recursive procedure that is used to solve the equilibrium industrial dynamics in stages 2 and 3, and is therefore the easiest to determine. Free entry with the standard product results in a zero-profit condition for traditional firms: $\Pi_{T,1} = \pi^{CN}\left(Z_T \middle| N_{T,1}, 0, 0, Z_i\right) - F_P = 0$. $N_{T,1}^*$ is the equilibrium number of traditional firms implied by this equation.

The solution to the model is thus made up of the SSE number of innovators, imitators and traditional firms in each stage, given the subjective expectations about and actual outcome of the uncertain innovation process: N_I^*, N_C^*, $N_{T,1}^*$, $N_{T,2}^*$, $N_{T,3}^*$. One can work back from this equilibrium investment and entry pattern to the outputs of each kind of firm and the price commanded by traditional and improved goods by substituting the equilibrium number

of firms into the Cournot-Nash equilibrium quantity $q^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)$ and price $P^{CN}\left(Z_j|N_T,N_I,N_C,Z_i\right)$. The equilibrium net profits of each type of firm can be evaluated in the same way. By construction, the net profits of traditional firms and imitators will always be equal to zero: $\Pi^*_{T,1}=\Pi^*_{T,2}=\Pi^*_{T,3}=\Pi^*_C=0$. This is not true of an innovator's *realised* profit, which very much depends on the outcome of the innovation process because the equilibrium number of innovators N^*_I is determined when firms are still enveloped in a 'cloud of uncertainty.' When this cloud dissipates, each will find itself with profits equal to $\Pi^*_I(Z_R)$, which may be greater than, equal to, or less than zero. In principle, one could derive a variety of other firm-level and industry-wide indicators, such as measures of industrial profitability, market power or consumer welfare. But, as Section 6.4 argues, these calculations call for a numerical methodology.

6.3 Analysis of the symmetric sustainable equilibrium

This section aims to establish some general properties of the specified model and to clarify its 'inner workings,' that is, to elucidate the mathematical and economic significance of the sequential structure of the model and the role of sunk costs in the determination of the SSE. In order to do so, it will be necessary to: (i) determine the necessary conditions for positive entry by traditional firms in stages 2 and 3, and for positive investment in imitative-R&D; (ii) derive the "sustainable entry" functions for imitating and traditional firms; (iii) solve for the equilibrium profit functions corresponding to traditional, imitating and innovating firms, and verify their central properties; and (iv) apply these results in a simple example that illustrates the nature of the SSE. The central result concerns the uniqueness of a SSE. In addition, the analysis reveals the exact mathematical relationship between the sustainable entry by different types of firms and the parameters pertaining to the cost and demand structure. One

feature that is revealed by this analysis is that, in limiting cases, small variations in the parameters may produce 'jumps' in the equilibrium numbers of innovating, imitating or traditional firms. It also shows that a wide range 'industry evolutions' can be supported in the present specification of the model, and it establishes precise conditions under which each may arise. Significantly, this finds that there are some endogenous limits to the evolutions that may arise in a SSE – specifically, the traditional and imitating firms sustainable entry functions in the third stage demonstrate that positive entry by imitators is incompatible with entry by traditional firms in that stage.

The claim about uniqueness can be stated thus: if a SSE with positive innovation $N_I^* > 0$ exists, it will be unique. A relatively informal proof of this claim runs as follows: (i) innovators' net profit functions $\Pi_{I,2}(Z_i)$ and $\Pi_{I,3}(Z_i)$ after anticipating the sustainable entry of traditional and imitating firm in each stage can be shown to be continuous and monotonically decreasing in N_I, for any admissible outcome of the innovation process $Z_i > Z_T$; therefore (ii) the cumulative discounted profit functions $\Pi_I(Z_i) = \Pi_{I,2}(Z_i) + \delta\,\Pi_{I,3}(Z_i)$ must be so too, because this property is preserved under summation; and, for the same reason, (iii) the expected profit function $\Pi_I = (1-p^*)\Pi_I(Z_L) + p^*\Pi_I(Z_H)$ is continuous and monotonically decreasing. Continuity and monotonicity guarantee that the maximum number of innovating firms N_I^* is unique, and we will make use of some diagrams to elucidate the intuition behind this result. Similarly, it is shown that the net profit functions for imitating firms $\Pi_C(Z_i)$ is continuous and monotonically decreasing in the number of innovators N_I for any $Z_i > Z_T$, as are traditional firms' net profits at stages 2 and 3, $\Pi_{T,2}(Z_i)$ and $\Pi_{T,3}(Z_i)$. Therefore, the uniqueness of N_I^*

guarantees that of N_C^*, $N_{T,2}^*$, and $N_{T,3}^*$. Finally, as $\Pi_{T,1}$ is continuous and strictly decreasing in $N_{T,1}$, $N_{T,1}^*$ also is unique.

It is important to make two remarks about this argument. First, that the claim to uniqueness is circumscribed to markets with two or more firms $N_T + N_I + N_C \geq 2$. This is because, as we mentioned before, the Cournot-Nash equilibrium price, output and profits are only valid under this assumption. Second, that the presence of flat segments and 'kinks' that will be seen to characterise the profit curves across stages 2 and 3 is irrelevant to our claim because the SSE corresponds to the *maximum* number of feasible innovating, imitating and traditional firms (i.e., the largest number that can secure zero profits or expected profits), not to the number that maximise each firm's or the industry's expected profits. In the latter case, the profit functions would need to be strictly concave and differentiable, but not so with the equilibrium concept that has been put forward.

We now proceed to derive each of the sustainable entry and net profit functions, to verify their properties, and discuss the relationship they hold with the parameters in the model. In accordance with the recursive solution procedure that has been outlined in the previous section, this analysis begins by characterising the sustainable entry of traditional firms in stage 3. Recall that net profits of a traditional firm in stage 3 equal:

$$\Pi_{T,3}(Z_i) = \pi^{CN}\left(Z_T \middle| N_{T,3}(Z_i), N_I, N_C(Z_i), Z_i\right) - F_P$$

Introducing the expression for Cournot Nash equilibrium gross profits, net profits equal:

$$\Pi_{T,3}(Z_i) = 2S\left[\frac{(1-c)((4-\sigma)Z_T - \sigma(N_I + N_C(Z_i))(Z_i - Z_T))}{(4-\sigma)(4+\sigma(N_{T,3}(Z_i) + N_I + N_C(Z_i) - 1))}\right]^2 - F_P$$

This profit function is decreasing in the outcome of the innovation process $(Z_i - Z_T)$ for any $Z_i > Z_T$ and in the number of firms that have access to the newly developed product lines, $N_I + N_C(Z_i)$. In order to derive the equilibrium locus $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$, it is necessary to establish the parameter subspace in which a positive number of firms can make zero net profits by producing the conventional product line, and that 'region' in which the rivalry from the improved product is so great that profits are negative for any positive but arbitrarily small number of traditional firms. The critical number of firms marketing the new product, denoted as $\overline{N}(Z_i)$, can be derived in two steps. First, by setting $\Pi_{T,3}(Z_i)$ equal to zero, and solving for $N_I + N_C(Z_i)$, which gives:

$$N_I + N_C(Z_i) = \frac{4-\sigma}{\sigma} \frac{\sqrt{2S}(1-c)Z_T - \sqrt{F_P}(4+(N_{T,3}(Z_i)-1)\sigma)}{\sqrt{2S}(1-c)(Z_i - Z_T) + \sqrt{F_P}(4-\sigma)}$$

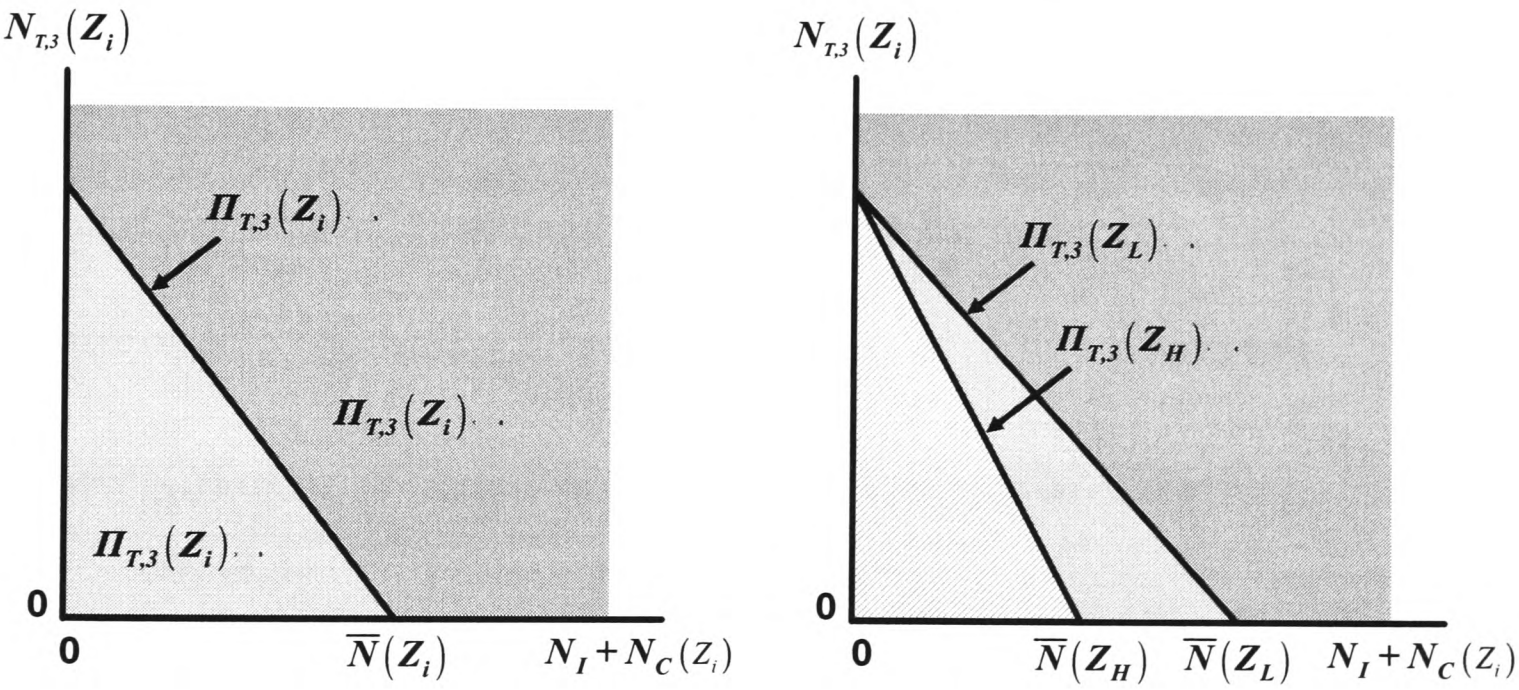
Second, suppose that the number of traditional firms in stage 3 is arbitrarily small. In the limit when the number of traditional firms goes to zero, the right part of the above expression becomes:

$$\overline{N}(Z_i) = \frac{4-\sigma}{\sigma} \frac{\sqrt{2S}(1-c)Z_T - \sqrt{F_P}(4-\sigma)}{\sqrt{2S}(1-c)(Z_i - Z_T) + \sqrt{F_P}(4-\sigma)}$$

If the number of innovating and imitating firms $N_I + N_C(Z_i)$ is greater than or equal to $\overline{N}(Z_i)$, then production of the conventional product is unprofitable for any positive number of traditional firms. Conversely, if $N_I + N_C(Z_i) < \overline{N}(Z_i)$, a positive number of traditional firms can be supported by the market making zero profits. To illustrate this diagrammatically, the left hand diagram in Figure 6.3 depicts a traditional firm's zero-profit iso-profit curve; in the area below it, a typical firm producing the conventional product would make a positive

net profit; above it, the firm registers a net loss. $\overline{N}(Z_i)$ can be found by moving down this iso-profit curve to the bottommost point, at which the number of traditional firms is zero. $\overline{N}(Z_i)$ is decreasing in $(Z_i - Z_T)$ because, the greater the relative difference between the improved product and the conventional one, the lower the number of innovators and imitators that make the conventional product line unprofitable to produce. In addition, $\overline{N}(Z_i)$ is decreasing in F_P since the profit of a traditional firm is decreasing in its fixed cost of production. The right hand side of Figure 6.3 illustrates the difference between the slopes taken by the iso-profit curve for the two anticipated values of $Z_i = Z_L, Z_H$.

Figure 6.3 Traditional firms zero-profit iso-profit curves



Having identified these regions, the number of traditional firms supported by the market in stage 3 (conditional on the rivalry from the new product line) can be derived by setting $\Pi_{T,3}(Z_i)$ equal to zero and solving for $N_{T,3}(Z_i)$.

$$N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i) =$$

$$\begin{cases} \sqrt{\frac{2S}{F_P}}\frac{(1-c)Z_T}{\sigma} - \frac{4-\sigma}{\sigma} - (N_I + N_C(Z_i))\left[1 + \sqrt{\frac{2S}{F_P}}\frac{(1-c)(Z_i - Z_T)}{4-\sigma}\right] & \text{if } N_I + N_C(Z_i) < \overline{N}(Z_i) \\ 0 & \text{if } N_I + N_C(Z_i) \geq \overline{N}(Z_i) \end{cases}$$

In the region in which a positive number of traditional firms is supported, $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ is decreasing in the marginal and fixed costs of production, c and F_P; and increasing in the market size S. The sign of the relationship between $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ and the degree of horizontal differentiation is ambiguous. The last term shows that there is a linear relationship between $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ and $N_I + N_C(Z_i)$ and, interestingly, that the entry by an additional firm producing the new product line leads to a more than proportional decrease in the number of traditional firms that the market can support. Conditional on remaining in the region $N_I + N_C(Z_i) < \overline{N}(Z_i)$, the derivative of $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ with respect to N_I or $N_C(Z_i)$ is equal to:

$$\frac{\partial N_{T,3}^{MAX}}{\partial N_I} = \frac{\partial N_{T,3}^{MAX}}{\partial N_C(Z_i)} = -1 - \sqrt{\frac{2S}{F_P}}\frac{(1-c)(Z_i - Z_T)}{4-\sigma}$$

The derivative shows that additional innovators or imitators would have a more pronounced effect when the improvement in relative quality is more significant, $\uparrow (Z_i - Z_T)$; or when there is less horizontal differentiation between the products, $\uparrow \sigma$.

The next step in the solution is to substitute the equilibrium locus $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ into the net profits of an imitator. Before this substitution, a typical imitator's net profits are simply:

$$\Pi_C(Z_i) = \pi^{CN}\left(Z_i \middle| N_{T,3}, N_I, N_C, Z_i\right) - F_P - F_C$$

$$= 2S\left[\frac{(1-c)\left((4-\sigma)Z_i + \sigma N_{T,3}(Z_i - Z_T)\right)}{(4-\sigma)\left(4+\sigma\left(N_{T,3}(Z_i) + N_I + N_C(Z_i) - 1\right)\right)}\right]^2 - F_P - F_C$$

if $N_I > 0$, as otherwise there is no innovation to clone. Substituting $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ into the equilibrium net profit function and simplifying gives:

$$\Pi_C(Z_i) =$$

$$\begin{cases} 2(1-c)(Z_i - Z_T)\left[\dfrac{\sqrt{2SF_P}(4-\sigma) + S(1-c)(Z_i - Z_T)}{(4-\sigma)^2}\right] - F_C & \text{if } N_I + N_C(Z_i) < \overline{N}(Z_i) \\ 2S\left[\dfrac{(1-c)Z_i}{4+\sigma(N_I + N_C(Z_i) - 1)}\right]^2 - F_P - F_C & \text{if } N_I + N_C(Z_i) \geq \overline{N}(Z_i) \end{cases}$$

Once the strategic effect on traditional firms' entry is taken into account, an imitator's profit schedule will be composed of two segments, one corresponding to the region in which the cloning firms face active competition from the traditional firms, and one where only the newer product line is marketed. In the former, an imitator's profits are independent of the number of innovating or imitating firms: competition from traditional firms leads to a flat segment in the net profit schedule $\Pi_C(Z_i)$. Let the net profits received by a typical imitator in the region $N_I + N_C(Z_i) < \overline{N}(Z_i)$ be denoted $\Pi_C^{MAX}(Z_i)$. This sum is decreasing in the sunk cost of imitation and the marginal cost; and increasing in the relative quality of the new product, the fixed cost of production, the size of the market and the inverse degree of horizontal differentiation. If $N_I + N_C(Z_i) \geq \overline{N}(Z_i)$, the profits of an imitating firm are increasing in the market size; and decreasing in the number of innovators and imitators $N_I + N_C(Z_i)$, the marginal cost, the fixed cost of production and the sunk cost of imitation.

The competitive pressure from the $N_{T.3}^{MAX}(N_I, N_C(Z_i), Z_i)$ firms marketing the conventional product in stage 3 does not create any discontinuities or other irregularities in imitators profit function. $\Pi_C(Z_i)$ is a continuous and monotonically decreasing function in $N_I + N_C(Z_i)$. This is an indispensable property if the maximum number of imitating firms receiving zero profits $N_C^{MAX}(N_I, Z_i)$ is to be unique given the number of innovators N_I.

Figure 6.4 Traditional firms' influence on imitators profits

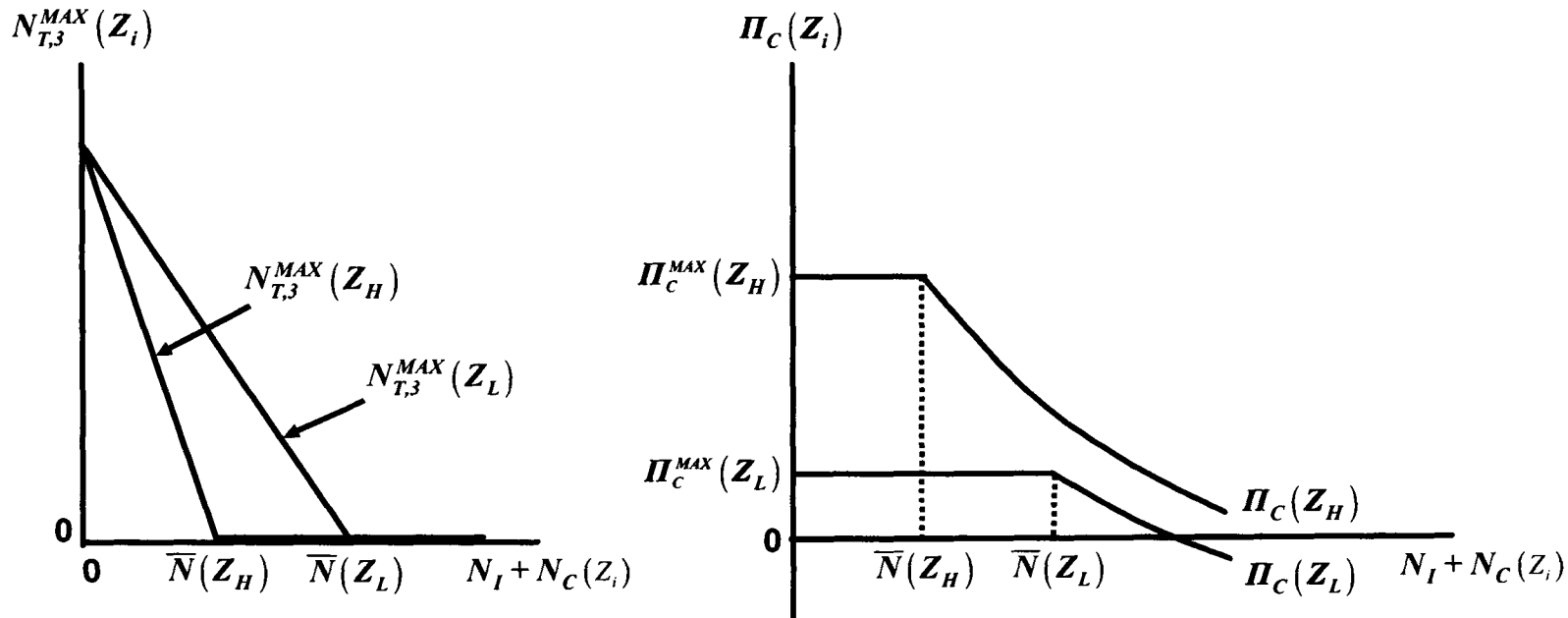


Figure 6.4 may serve to clarify what has been established thus far and, more importantly, to learn what implications exist for the equilibrium entry by imitating firms, the next step in this analysis. The diagram on the left shows the sustainable entry of traditional firms, conditional on the number of innovating and imitating firms, and the quality of the latter firms' product. This displays the negative, linear relationship between $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ and $N_I + N_C(Z_i)$. $N_{T,3}^{MAX}(N_I, N_C(Z_i), Z_i)$ drops to zero at $\overline{N}(Z_i)$ and is zero for any $N_I + N_C(Z_i) > \overline{N}(Z_i)$. It also shows that, given a certain $N_I + N_C(Z_i)$, the number of traditional firms that can be supported is lower with a major innovation $Z_i = Z_H$ than with a minor one $Z_i = Z_L$.

The right hand diagram in Figure 6.4 depicts an imitator's profit function once the effect of competition from the conventional product has been factored in. The central feature is the flat segment for $N_I + N_C(Z_i) < \overline{N}(Z_i)$, after which profits are decreasing in $N_I + N_C(Z_i)$. The flat segments in these curves have been drawn as being above zero, but there is no reason why this should be so. It is possible for $\Pi_C^{MAX}(Z_i)$ to be negative and, if this is the case, net profits will be negative for any $N_I + N_C(Z_i)$. This is simply to say that, given the sunk cost of imitation and the intensity of competition from the traditional firms, no imitative-R&D can be financed profitably, regardless of the number of innovators that are already marketing the new product. Indeed, a sufficient condition for imitation not to be undertaken in equilibrium is for the profits of a typical imitator to be negative for $N_I + N_C(Z_i) < \overline{N}(Z_i)$.

Looked upon from another angle, this gives the first of two necessary conditions for the equilibrium number of imitators to be positive when the innovation has quality Z_i, which is for $\Pi_C^{MAX}(Z_i) \geq 0$. The second necessary condition for $N_C^{MAX}(N_I, Z_i) > 0$ is for the number of innovating firms to be below the critical level that effectively blockades entry by imitators. Following the notation used earlier, this critical number is denoted $\overline{\overline{N}}(Z_i)$. It can be derived in a similar way to $\overline{N}(Z_i)$, by finding the number of innovating firms for which an imitator's profit is zero and letting the number of imitating firms tend to zero – *conditional* on there not being any traditional firms in the industry. This derivation gives:

$$\overline{\overline{N}}(Z_i) = \sqrt{\frac{2S}{F_P + F_C}} \frac{(1-c)Z_i}{\sigma} - \frac{4-\sigma}{\sigma}$$

Alternatively, this expression can be written as:

$$\overline{\overline{N}}(Z_i) = \frac{(4-\sigma)}{\sigma} \frac{\sqrt{2S}(1-c)Z_i - \sqrt{F_P + F_C}(4-\sigma)}{\sqrt{F_P + F_C}(4-\sigma)}$$

The number of innovators $\overline{\overline{N}}(Z_i)$ that automatically blockades subsequent entry by imitators is increasing in the market size, the quality of the new product; and decreasing in the marginal cost, the fixed cost of production and the sunk cost of imitation. Using $\overline{\overline{N}}(Z_i)$ and the other necessary conditions, it is possible to derive the sustainable entry function for imitating firms,

$$N_C^{MAX}(N_I, Z_i) =$$

$$\begin{cases} \sqrt{\dfrac{2S}{F_P + F_C}} \dfrac{(1-c)Z_i}{\sigma} - \dfrac{4-\sigma}{\sigma} - N_I & \text{if } 0 < N_I < \overline{\overline{N}}(Z_i) \text{ and } \Pi_C^{MAX}(Z_i) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

In the region in which $0 < N_I < \overline{\overline{N}}(Z_i)$ and $\Pi_C^{MAX}(Z_i) \geq 0$, there is a one-to-one trade-off (or 'crowding out') between the number of innovators and the sustainable number of imitators: the presence of one more innovating firm prevents one firm from imitating the improved product. Additionally, it is clear that the number of imitating firms that can be supported is increasing in the quality of the product they are copying and the market size, but decreasing in the fixed cost of production and imitation, and the marginal cost. The significance of the flat segment in the imitators profit function $\Pi_C(Z_i)$ is that the equilibrium investment in imitative-R&D $N_C^{MAX}(N_I, Z_i)$ may 'jump' from a positive amount to zero in response to an infinitesimally small variation in the parameters, instead of increasing in a smooth fashion. In

the limit case where $\Pi_C^{MAX}(Z_i)=0$, if there is a small increase in, say, the sunk cost of imitation F_C the number of imitating firms drops from $\overline{\overline{N}}(Z_i)-N_I$ to 0.

Substituting the equilibrium investment $N_C^{MAX}(N_I,Z_i)$ into the traditional firm's equilibrium entry in stage 3, it is possible to arrive at $N_{T,3}^{MAX}\left(N_I,N_C^{MAX}(N_I,Z_i),Z_i\right)$, a function for the sustainable entry by traditional firms that is a function of the number of innovating firms and the quality of the latter firms' product:

$$N_{T,3}^{MAX}\left(N_I,N_C^{MAX}(N_I,Z_i),Z_i\right)=$$

$$\begin{cases} \sqrt{\dfrac{2S}{F_P}}\dfrac{(1-c)Z_T}{\sigma}-\dfrac{4-\sigma}{\sigma}-N_I\left[1+\sqrt{\dfrac{2S}{F_P}}\dfrac{(1-c)(Z_i-Z_T)}{4-\sigma}\right] & \text{if } N_I<\overline{N}(Z_i) \text{ and} \\ & \left(\Pi_C^{MAX}(Z_i)<0 \text{ or } N_I\geq\overline{\overline{N}}(Z_i)\right) \\ 0 & \text{otherwise} \end{cases}$$

This indicates that firms will market the conventional product line in stage 3 if (i) it is unprofitable for a positive number of firms to imitate the product and (ii) the number of innovating firms does not surpass $\overline{N}(Z_i)$. Notice that a small change in the parameters that causes the equilibrium investment in imitative-R&D to fall to zero may generate a discrete upward change or 'jump' in the feasible number of traditional firms in stage 3.

If $N_{T,3}^{MAX}\left(N_I,N_C^{MAX}(N_I,Z_i),Z_i\right)$ and $N_C^{MAX}(N_I,Z_i)$ are substituted into the net profit function for a typical innovator in stage 3,

$$\Pi_{I,3}(Z_i)=\pi^{CN}\left(Z_i\middle|N_{T,3}^{MAX}\left(N_I,N_C^{MAX}(N_I,Z_i),Z_i\right),N_I,N_C^{MAX}(N_I,Z_i),Z_i\right)-F_P$$

one obtains, after some algebraic manipulations, an expression for an innovator's profits that has factored in all the anticipated competition from traditional and imitating firms:

$$\Pi_{I,3}(Z_i) =$$

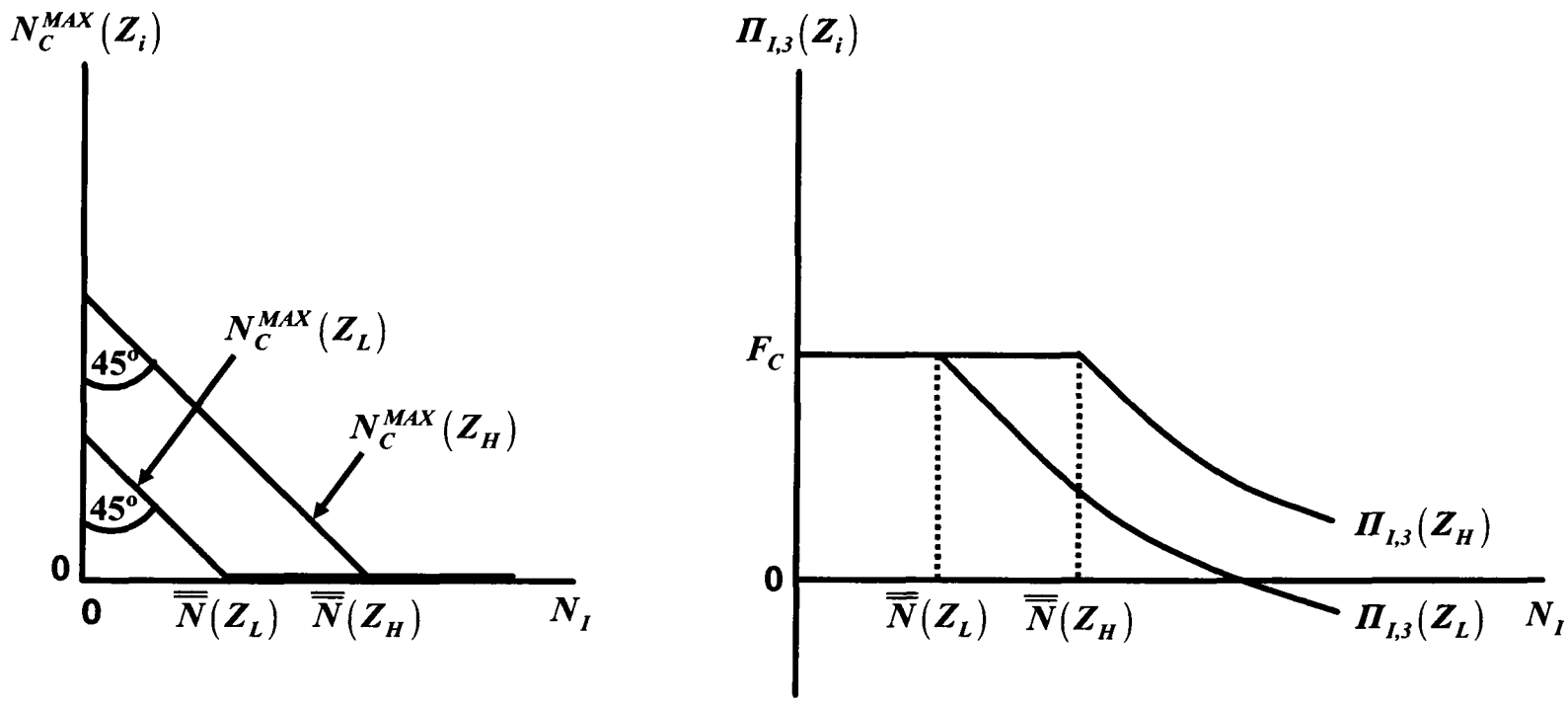
$$\begin{cases} F_C & \text{if } N_I < \overline{\overline{N}}(Z_i) \text{ and } \Pi_C^{MAX}(Z_i) \geq 0 \\ 2(1-c)(Z_i - Z_T)\left[\dfrac{\sqrt{2SF_P}(4-\sigma)+S(1-c)(Z_i-Z_T)}{(4-\sigma)^2}\right] & \text{if } N_I < \overline{N}(Z_i) \text{ and } \left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \geq \overline{\overline{N}}(Z_i)\right) \\ 2S\left[\dfrac{(1-c)Z_i}{4+\sigma(N_I-1)}\right]^2 - F_P & \text{otherwise} \end{cases}$$

The intuition behind $\Pi_{I,3}(Z_i)$ is the following. If conditions allow imitating firms to enter, innovators receive a quasi-rent in stage 3 equal to the cost of cloning the new product F_C, as they no longer need to disburse any resources to market this – the sunk cost of innovation F_I is completely accounted for in the second stage. If no imitators can profitably enter, but the conventional product is marketed, there will still be a flat segment in the profit function, corresponding to the rent for owning a better product design. Lastly, if entry by traditional and imitating firms is blockaded, each innovator's profits are decreasing in the number of innovators. Because entry by traditional and imitating firms at stage 3 are mutually exclusive, the expression for innovators profits at the third stage shares two decisive properties with imitating firms profit function, namely $\Pi_{I,3}(Z_i)$ is continuous and monotonically decreasing in N_I, for any positive number of innovators.

Figure 6.5 illustrates the case in which imitators find it profitable to clone both minor and major innovations, the same that was portrayed in Figure 6.4. The left hand diagram traces the sustainable entry function corresponding to imitating firms. This shows the one-for-one displacement of imitators provoked by prior investments in innovation, the critical $\overline{\overline{N}}(Z_i)$ that blockades entry for the forecasted outcomes $Z_i = Z_H, Z_L$, and that (for a given

N_I) more imitators can be supported when a major innovation is available for cloning. The right hand figure plots the profit functions of a typical innovator in stage 3. In this case, if the number of innovators lies in $0 < N_I < \overline{\overline{N}}(Z_L)$, each will receive as a rent F_C, regardless of the outcome of innovation. When N_I lies in the interval $\left[\overline{\overline{N}}(Z_L), \overline{\overline{N}}(Z_H)\right]$, they receive F_C with a major innovation; with a minor innovation they get a lower figure, one that is decreasing in N_I and not necessarily positive. If the number of innovators exceeds $\overline{\overline{N}}(Z_H)$, their profits will be less than F_C and decreasing in N_I for both Z_L and Z_H.

Figure 6.5 Imitators' influence on innovators profits in stage 3



By following the same steps that have been outlined, it is possible to find the sustainable number of traditional firms in the second stage $N_{T,2}^{MAX}(N_I, Z_i)$, and an expression for an innovator's profits that has taken into account the 'competitive pressure' from firms marketing the conventional product. The sustainable entry of traditional firms can be shown to be:

$$N_{T,2}^{MAX}(N_I, Z_i) =$$

$$\begin{cases} \sqrt{\frac{2S}{F_P}}\frac{(1-c)Z_T}{\sigma} - \frac{4-\sigma}{\sigma} - N_I\left[1+\sqrt{\frac{2S}{F_P}}\frac{(1-c)(Z_i-Z_T)}{4-\sigma}\right] & \text{if } N_I < \overline{N}(Z_i) \\ 0 & \text{if } N_I \geq \overline{N}(Z_i) \end{cases}$$

and by substituting this function into the profits earned by an innovator $\Pi_{I,2}(Z_i)$ in stage 2,

$$\Pi_{I,2}(Z_i) = \pi^{CN}\left(Z_i \middle| N_{T,2}^{MAX}(N_I, Z_i), N_I, 0, Z_i\right) - F_P - F_I$$

it is possible to get the following expression for an innovator's profits:

$$\Pi_{I,2}(Z_i) =$$

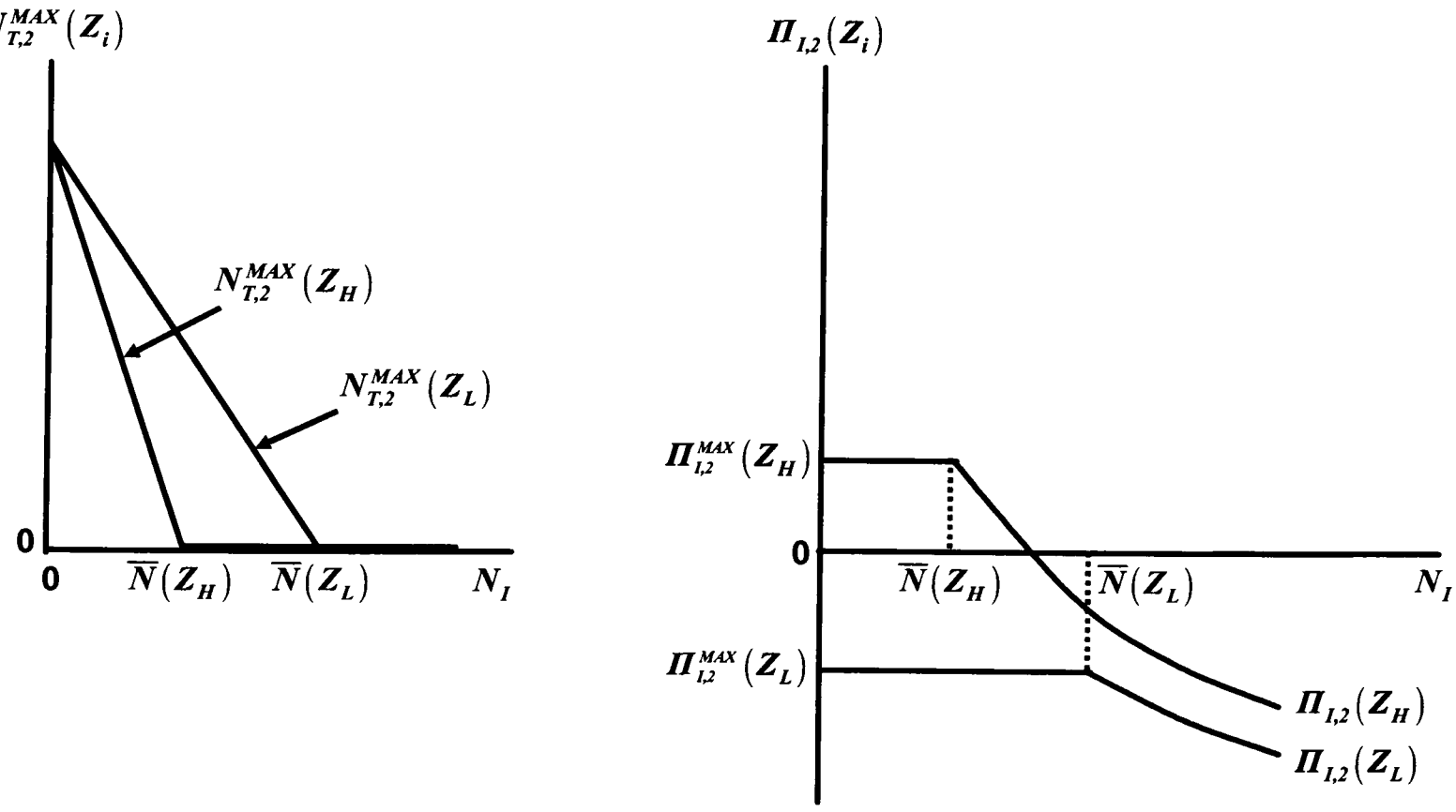
$$\begin{cases} 2(1-c)(Z_i - Z_T)\left[\frac{\sqrt{2SF_P}(4-\sigma)+S(1-c)(Z_i-Z_T)}{(4-\sigma)^2}\right] - F_I & \text{if } 0 < N_I < \overline{N}(Z_i) \\ 2S\left[\frac{(1-c)Z_i}{4+\sigma(N_I-1)}\right]^2 - F_I - F_P & \text{if } N_I \geq \overline{N}(Z_i) \end{cases}$$

$\Pi_{I,2}(Z_i)$ are the profits that an innovating firm could receive in the space of time before investments in imitative-R&D deliver a marketable copy of their product design that brings about a decline in their revenue. An innovator's second stage profit curve has the same profile as that found for an imitating firm. Once more, notice that $\Pi_{I,2}(Z_i)$ is continuous and monotonically decreasing in N_I for any positive number of innovators.

Let $\Pi_{I,2}^{MAX}(Z_i)$ denote the value taken in the flat segment corresponding to positive entry by traditional firms. Then, the right hand diagram in Figure 6.6 shows one possible outcome, in which $\Pi_{I,2}^{MAX}(Z_H) > 0$ and $\Pi_{I,2}^{MAX}(Z_L) < 0$. If that were the case, $0 < N_I < \overline{N}(Z_H)$

innovators could recoup the totality of their investment F_I and make a profit in stage 2 with a major innovation; if there were more innovators, however, the 'product lead-time effect' may be insufficient for them to break-even. When a minor innovation has been developed, each would sustain a loss of at least $\Pi_{I,2}^{MAX}(Z_L)$ in stage 2, whatever the number of innovators. Naturally, innovating firms may experience a loss with both a minor and a major innovation if F_I is large enough to make $\Pi_{I,2}^{MAX}(Z_H)<0$ for $Z_i = Z_H, Z_L$. Conversely, they may break-even early with both a minor and a major improvement in product design if F_I is sufficiently small, and hence $\Pi_{I,2}^{MAX}(Z_i)\geq 0$ for $Z_i = Z_H, Z_L$.

Figure 6.6 Traditional firms' influence on innovators profits in stage 2



Lastly, the sustainable entry by traditional firms in stage 1, which plays no part in the recursive solution, can readily be determined. Traditional firms' net profits at this stage are:

$$\Pi_{T,1} = \pi^{CN}\left(Z_T \middle| N_{T,1}, 0, 0, Z_i\right) - F_P$$

$$= 2S\left[\frac{(1-c)Z_T}{4+\sigma(N_{T,1}-1)}\right]^2 - F_P$$

By setting this expression equal to zero and solving for $N_{T,1}$, one obtains the equilibrium number of traditional firms in stage 1:

$$N_{T,1}^* = \sqrt{\frac{2S}{F_P}}\frac{(1-c)Z_T}{\sigma} - \frac{4-\sigma}{\sigma}$$

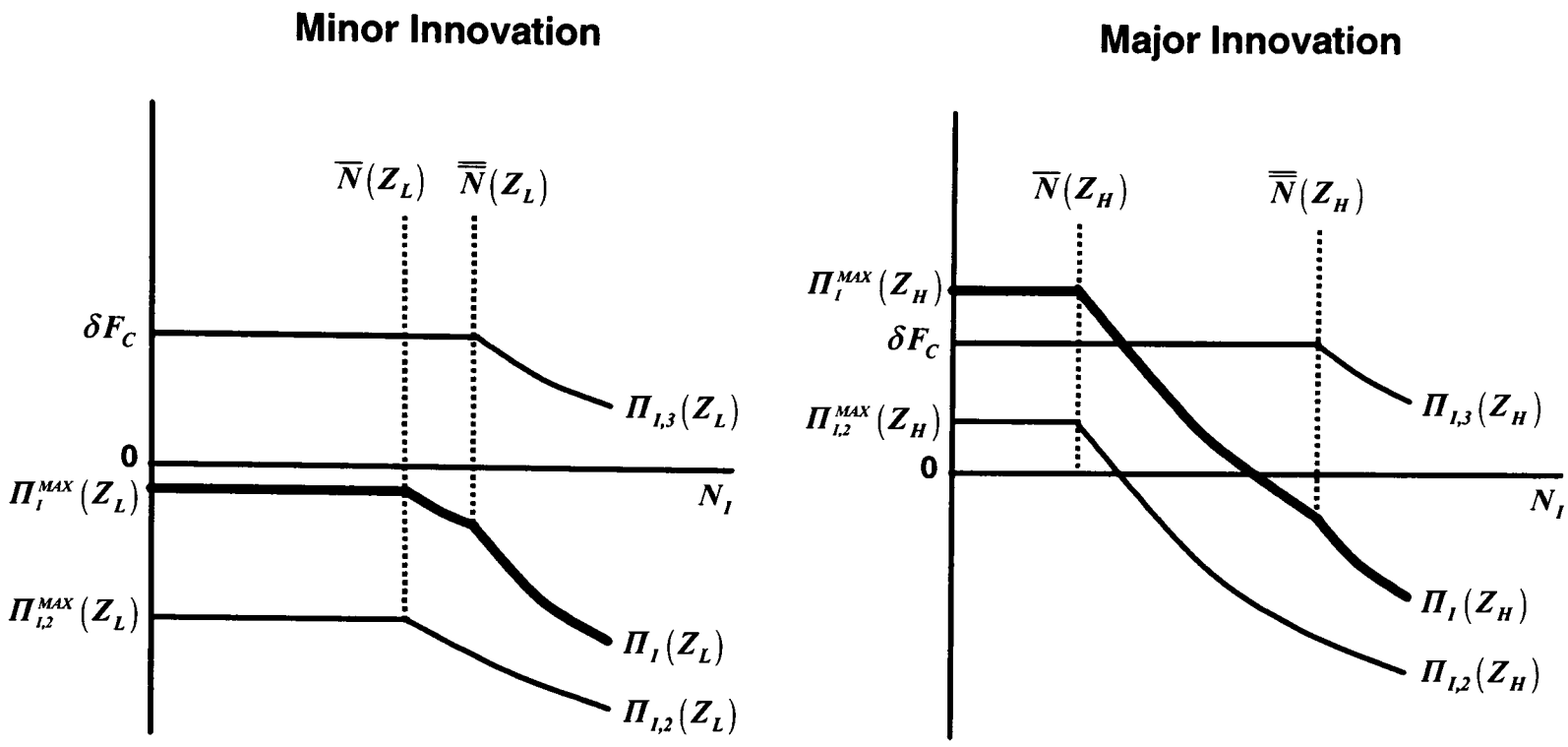
By drawing on the innovators' profit functions in stages 2 and 3, one could derive the cumulative, discounted profits for any outcome of the innovation process, and ultimately the innovators' expected profits. All the same, the foregoing analysis of the innovators profit functions indicates that a sustainable equilibrium with a positive number of firms investing in innovative-R&D, $N_I^* > 0$, cannot be guaranteed without (overly) restricting the admissible parameter space. Nor can an exact solution for N_I^* be derived without additional constraints on the values taken by the parameters determining the demand and cost structures. There is a wide range of possible 'industry evolutions' in the present specification of the model and, as the investment in innovative-R&D is taken under uncertainty and in advance, innovators expected profit function will be considerably more intricate than the profit functions of traditional or imitating firms. But, crucially, the expected profit function will be continuous and monotonically decreasing in N_I, because of the arguments given at the start of the section; namely, that the sum of continuous and monotonically decreasing functions is a continuous and monotonically decreasing function. This is the key result of this chapter, as it guarantees the uniqueness of the SSE for $N_I > 0$. It is particularly important if one is to have confidence in the numerical analysis of the specified problem undertaken in Chapter 7.

The precise shape of the expected profit curve depends on whether, for a given number of innovators, it is profitable to market the conventional product at a given stage, and whether imitation is feasible or not, for any outcome of the innovation process. To be more precise, the expressions $\Pi_{I,2}(Z_i)$ and $\Pi_{I,3}(Z_i)$ show that the key to unravelling the complexities of the expected profit function is given by: $\overline{N}(Z_i)$, the number of firms with access to the new product line (innovators and/or imitators) that would automatically blockade traditional firms' entry; $\overline{\overline{N}}(Z_i)$, the number of innovating firms blockading imitating firms' entry; and $\Pi_C^{MAX}(Z_i)$, the profit made by each imitator when entry is not blockaded (i.e., if $N_I < \overline{\overline{N}}(Z_i)$). By evaluating the ordering of these critical values, the evolution of the industry can be exactly determined for any given number of innovating firms and any given outcome of the innovation process. This would allow one to ascertain the 'binding' expression for expected profits, that is, the one that is relevant given the structural conditions implied by the parameters c, F_P, F_I, etc.

Restricting the parameter subspace to prove the existence of the equilibrium having positive investment in innovation, or to get an exact solution for the equilibrium number of innovators, does not appear to be a fruitful line of analysis. Indeed, the array of equilibria that be supported in the model can be viewed as one of its strengths. Rather than restricting the analysis to a particular case, a better course of action is to apply what has been learnt so far in a diagrammatic analysis of a concrete example, as this serves to illustrate the central mechanisms at play in the determination of the SSE. Suppose that the critical values for $\overline{N}(Z_i)$ and $\overline{\overline{N}}(Z_i)$ are ordered as follows, $\overline{N}(Z_H) < \overline{N}(Z_L) < \overline{\overline{N}}(Z_L) < \overline{\overline{N}}(Z_H)$, and that $\Pi_C^{MAX}(Z_i) \geq 0$ for $Z_i = Z_H, Z_L$. This ordering implies that the anticipated industrial evolutions are the ones exemplified by the previous figures: there is imitation of a minor

innovation so long as the number of innovators is positive, but does not exceed $\overline{\overline{N}}(Z_L)$, and of a major innovation when N_I is less than $\overline{\overline{N}}(Z_H)$.

Figure 6.7 Innovators net profits

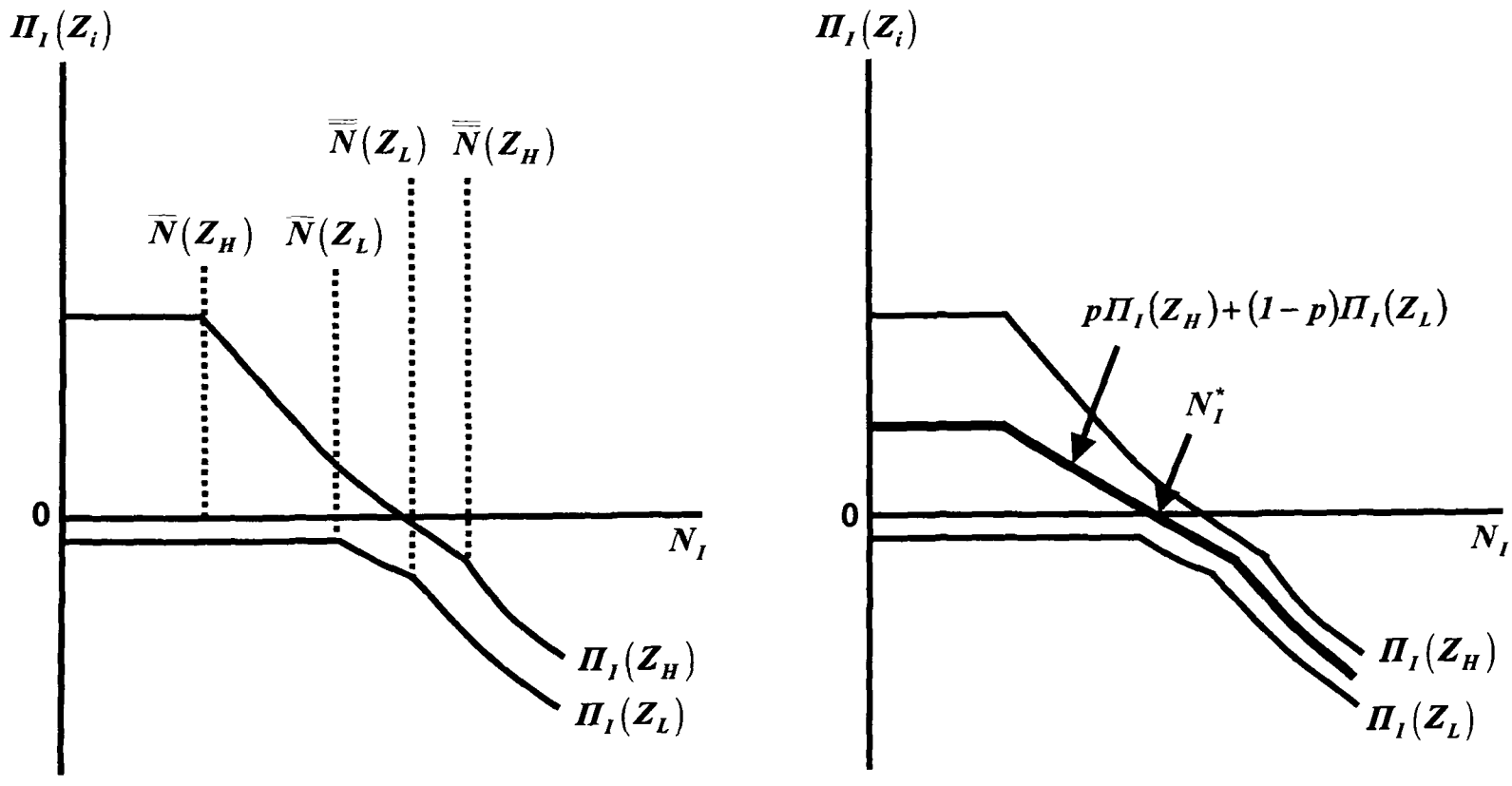


Under these conditions, the cumulative and discounted profit function for each Z_i, $\Pi_I(Z_i)=\Pi_{I,2}(Z_i)+\delta\,\Pi_{I,3}(Z_i)$, could be as depicted in Figure 6.7. The discounted sum of profits has a similar profile in each outcome of the innovation process: $\Pi_I(Z_i)$ has a flat segment in the region $0<N_I<\overline{N}(Z_i)$, it will be decreasing in the number of innovators for $N_I \geq \overline{N}(Z_i)$, and the curve will become steeper after the kink at $\overline{\overline{N}}(Z_i)$. These three segments in $\Pi_I(Z_i)$ correspond to well-defined evolutions in the industry: if the number of innovating firms is relatively small $0<N_I<\overline{N}(Z_i)$, the conventional product line will be marketed in the second stage but will be displaced by clones of the new product in stage 3; if the number of innovators rises to the values in the second segment $\overline{N}(Z_i)\leq N_I<\overline{\overline{N}}(Z_i)$, the conventional product is not worth producing in stage 2, but clones will nonetheless be

marketed in stage 3; if the investment in own-product development is even larger $\overline{\overline{N}}(Z_i) < N_I$, the market will not support any traditional or imitating firms.

To consider what this implies for expected profits and hence the equilibrium number of innovators, Figure 6.8 redraws the innovators' profit curves in each of the two outcomes on the left diagram, marking the position of the 'kinks' in each curve, and adds to these a third line on the right hand side diagram. This thick line represents the expected profit function of a typical innovator, $\Pi_I = (1-p^*)\Pi_I(Z_L) + p^*\Pi_I(Z_H)$. This is a weighted sum of the profit curves, where $1-p^*$ and p^* are the firms' common, subjective probabilities of a minor and a major innovation respectively.

Figure 6.8 Innovators expected profits



In this example, expected profits curve will have a flat segment when there are between zero and $\overline{N}(Z_L)$ innovators, and decrease in N_I when the number of innovators exceeds $\overline{N}(Z_L)$, for any p^* in the interval $[0,\ 1]$. The steepness of this negative slope depends on whether the number of innovators is above or below each of the successive 'kinks' that determine the

critical aggregate scale of investment in innovative-R&D that blockades either traditional or imitating firm in each state. N_I^* is determined by moving down the expected profit curve in Figure 6.8 to the point at which it crosses the zero axis. This is the maximum number of innovating firms that can expect to make zero profits. Clearly, if the flat segment had exactly coincided with the zero axis, the equilibrium number of innovators would be the rightmost point in the segment.

In order to completely characterise the equilibrium investments in imitative-R&D and the entry by traditional firms over the evolution of the industry, it would be necessary to substitute N_I^* and Z_R into $N_C^{MAX}(N_I, Z_i)$, $N_{T,3}^{MAX}(N_I, N_C^{MAX}(N_I, Z_i), Z_i)$ and $N_{T,2}^{MAX}(N_I, Z_i)$. If one supposes that, by mere chance, the realised product design Z_R coincides with either of the two anticipated states of nature $Z_i = Z_L, Z_H$, the outcome of the substitution is clear enough from Figure 6.8. Since N_I^* is greater than $\overline{N}(Z_L)$, which in turn is greater than $\overline{N}(Z_H)$, the conventional product will not be marketed in stages 2 or 3, for either outcome of the innovation process $Z_i = Z_L, Z_H$. And, as N_I^* is less than $\overline{\overline{N}}(Z_L)$, which itself is lower than $\overline{\overline{N}}(Z_H)$, there will be positive investment in imitative-R&D and thus clones of both a minor and a major innovation. N_I^* also pins down the actual profits that a typical innovator will receive in the two alternative evolutions (recall that, by assumption, any traditional and imitating firms obtain zero profits). The total discounted profits can be read off the figure: they are negative with a minor innovation Z_L and positive with a major innovation Z_H.

Another insight emerging from Figure 6.8 concerns the central role played by the subjective attitudes to technological uncertainty in the determination of N_I^* and, through this, on the viability of the conventional product line (in the second and third stages) and on the equilibrium investment in imitative-R&D. In states of 'systemic optimistic, ' in which firms

believe $p^* = \overline{p}$ and $1 - p^* = 1 - \overline{p}$, the expected profit curve will be an upward displacement (although not a parallel one) of that when firms are pessimistic, and they believe that $p^* = \underline{p}$ and $1 - p^* = 1 - \underline{p}$, as $\overline{p} > \underline{p}$. One can extrapolate from this that the equilibrium number of innovating firms N_I^* supported by optimistic beliefs (if it is positive) will be greater than that supported by pessimistic ones.

Depending on where exactly the equilibrium number of innovators falls and the realised quality of their product innovation, the industry evolution may be quite different. There is, however, an upper bound to N_I^*, given by the number of innovators that can make zero expected profits when they all expect to get a major innovation with certainty, $p^* = 1$. This means that the subjective probability has an important influence on the sustainable level of investment in innovative-R&D and on the subsequent evolution of the industry, but not an unbounded one. The following chapter will indicate what other implications may follow from different subjective estimates about the innovation process, in terms of industry concentration, price mark-ups, profit margins, sales volumes and consumer welfare.

In this example, if p^* is close enough to zero (and hence $1 - p^*$ is close enough to one), N_I^* would drop to zero. And the use of the word 'drop' to describe this is not fortuitous: there could be a discrete change in the SSE number of innovators from $\overline{N}(Z_H)$ to 0 in response to an infinitesimal downward variation in p^*. In fact, this feature of the expected profits function implies that N_I^* can never lie in the open interval $\left(0, \overline{N}(Z_H)\right)$. Because $\Pi_{I,2}(Z_i)$ and $\Pi_{I,3}(Z_i)$ have a flat segment for any $Z_i > Z_T$, the expected profit curve necessarily has one too, which suggests that the possibility that N_I^* responds abruptly to small variations in the parameters will be a general feature of the specified model. But this only occurs in limiting cases, where N_I^* is 'very close' to the critical number of innovators at

which there is the first 'kink' in expected profits, the value of which will depend on the parameter values and can be very small. This is one of the features that will generally be present in a SSE with $N_I^* > 0$. Others are likely to depend on the parameter values determining the exact ordering of $\overline{N}(Z_i)$ and $\overline{\overline{N}}(Z_i)$, and the value taken by $\Pi_C^{MAX}(Z_i)$.

6.4 Conclusions

This chapter introduced a game theoretic approach to study the evolving industrial structure arising from successive waves of innovation and imitation. This features two methodological novelties: the explicit analysis of fundamental uncertainty surrounding technological advances, in which firms' profit expectations and investment decisions depend to some extent on subjective evaluation of opportunities that cannot be reduced to empirical risk assessments; and the endogenous determination of the market structure and R&D expenditures in a dynamic setting, in which details of the sequence of entry decisions matters.

With the object of solving the specified multistage model to which this approach leads, we put forward an equilibrium concept, denominated "symmetric sustainable equilibrium," and formalised a recursive solution procedure by which this equilibrium could be found based on zero profit and expected profit conditions. A formal analysis was then able to prove that: (i) if a SSE with positive investment in innovation existed, it would be unique; (ii) a wide range of possible "industry evolutions" could be supported, including surges of "innovative entry" and ensuing episodes of "creative destruction," in which traditional producers are driven from the industry and innovators rents' are eventually eroded; (iii) there are exact conditions under which certain other "historical evolutions" can be shown to be infeasible; (iv) in limiting cases, small variations in the model's parameters may produce 'jumps' in the equilibrium numbers of innovating, imitating or traditional firms.

A variety of assumptions, some of them quite restrictive, were made to keep the problem tractable. These concern the cost and demand structure, the potential range of outcomes from research directed towards innovation and imitation, the nature of the firms' beliefs regarding technological uncertainty and, finally, the sequence of entry decisions. With respect to the cost structure, the model assumed that firms faced symmetric production costs. This carries the implication that production costs were taken to be independent of the product design and the experience of the firm in question. In actual fact, firms typically need to adapt their manufacturing processes to produce the new products, which would increase the fixed cost of innovators or imitators vis-à-vis traditional producers and in all probability alter their marginal costs as well. While one can readily suppose that the additional fixed costs of production incurred by product development are already included in the sunk costs of innovation and imitation, extending the model to a heterogeneous marginal cost would prove problematic, given the intimate connection of this cost to the calculation of the Cournot-Nash equilibrium outputs. Of course, the current product design is only one factor impinging on a firm's cost structure; if firms moved down a learning curve as they accumulated production experience, the incentive to invest in product development would need to become a function of their entire production history.

The second limitation refers to the utility function that was used to derive the Cournot-Nash equilibrium profits. When consumer preferences are represented by a quality-augmented quadratic utility function, the firms face a linear demand system, which can be used to derive closed-form profit functions having several attractive features. Chief among these is that profits are increasing in firm's own product quality and decreasing in the quality of the products manufactured by rival firms, as noted by Symeonidis (2003). Since it is not at all obvious that other specifications of the utility function would yield the same results, a useful test of the robustness of the model would be to go through the derivations outlined in

the solution procedure with the Cournot-Nash profits corresponding to a Cobb-Douglas or a Constant Elasticity of Substitution (CES) utility function, for example. As closed-form solutions exist in these cases as well, this is certainly feasible. In connection to this, it may prove interesting to derive the SSE when firms compete in prices instead of quantities. The introduction of price competition for the linear demand structure is a straightforward exercise and, given the similarities between firms' equilibrium profit functions in the two settings, one would expect to find a solution that is almost identical to that derived in this chapter. Chapter 8 will employ numerical methods to solve for this 'type of competition'; the results obtained turn out to be very similar. Other authors have worked out the Bertrand equilibrium prices and quantities for the Cobb-Douglas and CES functional forms, and therefore the solution could be adapted to study price competition when representative consumers have different preferences too.[9]

With the intention of getting precise results about the effects of technological uncertainty on industrial dynamics, two assumptions were made concerning the potential "states of nature" that firms investing in innovation expected to find themselves in and the states that could actually come about. The first of these is that the research efforts geared towards new product development was expected to result in one of two alternative outcomes, a minor or a major innovation. This assumption was made to simplify our notation and description of the specified model and is by no means essential to the argument. Innovating firms' equilibrium entry (and consequently traditional and imitating firms' entry) would also be unique if one elaborated the model to allow for many more alternative outcomes, so long

[9] James Markusen, Richard Baldwin and Gianmacro Ottaviano have investigated the Cournot-Nash equilibrium corresponding to CES utility functions in several papers, and the central results are succinctly summarized by Head and Mayer (2001). The Bertrand equilibrium when representative consumers have CES utility functions was analysed in Dixit and Stiglitz (1977) in a setting of monopolistic competition, and other authors have extended their results to a setting when the number of firms is 'smaller.' Since the Cobb-Douglas function is a special of the Constant Elasticity of Substitution function, the equilibrium in that case can be obtained directly from the results for the CES. Symeonidis (2003) presents the Bertrand equilibrium for the quality-augmented quadratic utility functions.

as each of the admissible states results in some kind of innovation $Z_i > Z_T$, even if it is the most insignificant one can imagine. The reason should be clear from the proof supplied above: as a typical innovator's cumulative discounted profit functions was shown to be continuous and monotonically decreasing in N_I for any $Z_i > Z_T$, the expected profit function will be so too, no matter how many of these functions one adds together (using the appropriate probability weights). The only additional consideration one must make if one sought to enlarge the admissible state space turns on the meaning attached to what have been called states of systemic optimism or pessimism, but this should not prove a serious difficulty.

Much more significantly, we assumed that the outcome of the innovation process was perfectly correlated across innovating firms.[10] A natural extension would be to consider a richer set of states, in which more complex industrial configurations could arise – i.e., where some of the firms investing in innovative-R&D were more successful than others. There is reason to believe that relaxing this assumption would make the solution process a lot more complicated, as one would have to calculate the innovator's expected profits across a much larger number of possible outcomes. All the same, the added complexity of the solution procedure may not be as serious a problem as the other thorny issues that need to be resolved if this feature were to be amended in the model. If each firm can end up with a different product design, the probability of a particular industrial configuration may not be invariant to aggregate innovation expenditures, for instance. It also begs the question of how imitators select the product they will copy, and whether their entry can drive from the market lower quality variants developed by unsuccessful innovating firms. Instead of simply taking the number of innovators as given, it would be necessary to estimate how many of the firms

[10] This may be partially justified if the uncertainty surrounding original product development is predominantly conditioned by product- or industry-specific factors rather than by firm-specific differences (e.g., in the quality of a firm's research). But even then, a positive correlation of R&D outcomes is to be expected, not a perfect one.

developing a minor innovation would be supplanted by imitators copying the major innovations developed by others. In this setting, the exit from the market needs to include the departure by unsuccessful innovators, in addition to the shakeout of traditional producers analysed thus far, which may have significant implications for the nature of the equilibria that can be supported.

This chapter has argued that R&D decisions may be based on businessmen's gut feelings about the chances of a project's success when the true probabilities governing the process of innovation are unknown. More precisely, the decision process was represented by a subjective state-contingent pay-off matrix, in which the pay-off to alternative strategies depended upon the subsequent evolution of firm behaviours once the fundamental uncertainty is dispelled. For the sake of simplicity, we modelled this by assuming that the firms shared common beliefs about the two states that could ensue. Under strong symmetry assumptions about beliefs and outcomes, the inclusion of the Knightian uncertainty apparatus is open to criticism because almost identical results would emerge in a model following the standard risk approach. Although the rule of parsimony would suggest dispensing with the extra assumptions and notation, several reasons are in favour of keeping them. It allows the equilibrium concept to be distinguished from the perfect foresight equilibrium or others in which "agents' expectations turn out to be correct ex post." In the model, the equilibrium scale of R&D and eventual deviations from zero-profits are functions of the subjective point-estimates used by firms in their calculations of profitability and the final outcome from the innovation process. Importantly, this outcome can take on values that firms' believed had zero probability of occurring ex ante. As this chapter does not present a closed-form solution of the SSE, it is hard to identify the divergence between anticipated (subjective) and realised innovation outcomes and its role in the dynamic paths followed by the industry. The numerical simulations in the following chapter should make this point clearer.

The Knightian uncertainty framework is ideal for modelling differences in beliefs between agents, an alternative that is not explicitly detailed in the thesis, but which constitutes one of the most interesting and challenging avenues for future research. Firms may have diverse expectations, formed on the basis of different information about a specific research domain or attitudes to uncertainty, for example. To deal with this, one could introduce the process of belief formation directly into the model. One way of doing this would be by adding an extra stage to the current dynamic structure where the investments in R&D occur sequentially. The intuition emerging from game theoretic analyses in which the sequential entry by agents reveals their private information to the rest of the potential participants is that under certain conditions a common belief structure can still emerge; see, for example, Minehart and Scotchmer (1999).

A final comment in relation to the model's belief structure is that, by assuming the existence of a perfect capital market, the model has implicitly supposed that creditors share the same beliefs about the potential returns of the innovation projects as would-be innovators. This is crucial to the model because it allowed an unlimited number of entrants to borrow the fixed innovation costs. But in the context of Knightian uncertainty this is not so easy to justify, as the usual risk differentiation may be compounded by a "surge towards liquidity protection," in which irreversible investments are heavily discounted by investors.[11] If

[11] As Alan Greenspan remarked in a speech:

> "The surge toward less risky assets reflected dramatic increases in uncertainty, but still a risk differentiation judgment among various assets. The surge toward liquidity protection, however, is a step beyond, since it implies that any commitment is perceived as so tentative that the ability to easily reverse the decision is accorded a high premium. Risk differentiation, despite its recent abruptness, is, of course, a straight-forward feature of well-functioning capital markets. The enhanced demand for liquidity protection, however, reflected a markedly decreased willingness to deal with uncertainty--that is a tendency to disengage from risk-taking to a highly unusual degree. It is, of course, plausible that the current episode of investor fright will dissipate, and yield spreads and liquidity premiums will soon fall into more normal ranges. Indeed we are already seeing significant signs of some reversals. But that leaves unanswered the question of why such episodes erupted in the first place."

Remarks by Chairman Alan Greenspan, *The structure of the international financial system*, Annual Meeting of the Securities Industry Association, Boca Raton, Florida, November 5, 1998.

creditors have less information than firms about the prospects of innovating, they may be wary of accepting the assessments made by the firms; they may (rightly) view a high appraisal as a sign of an unfounded degree of optimism by managers and scientists. Even if their information were the same, creditors may be disinclined to finance uncertain investment projects. Whatever the motive for this divergence of opinion, it is likely to create a shortage of financing, which would limit the investments in innovative-R&D – and particularly for entrants that cannot rely on past cash flows to finance the sunk costs required.[12] Given the proposed frameworks' emphasis on the immeasurable uncertainty surrounding the technological innovation process, it may be sensible to introduce this kind of divergences in perceptions to discern whether financial limits curtail the range of equilibrium industry evolutions that can be supported.

The last set of assumptions concern the entry sequence in the model. Section 6.3 remarked that as R&D investments occur sequentially in the model, its structure bears some resemblance to a Stackelberg game with innovators as leaders and imitators as followers; and that, seen from this perspective, traditional producers could be conceived as an extra tier of followers that at each stage in the game enter the market, taking as given the number of firms that previously invested in some variety of product development. This dynamic structure is central to the recursive solution procedure that has been formalised, and hence to the results derived in this chapter. If one assumed, instead, that incumbent 'traditional firms' paid their fixed cost of production prior to the investments in innovation or imitation rather than in each stage, with the view of staying in the industry for the entire evolution, it is quite possible that subsequent R&D investments could be deterred.[13] This is certainly an interesting possibility,

[12] See, e.g., Rigotti (2003), who solves the optimal financial contract between a "cautious investor" and a "decisive entrepreneur," finding stark contrasts in the collateral demanded depending on the differential attitude towards uncertainty of the two agents.

[13] I would like to thank Dr. Margaret Meyer for pointing this out, thus encouraging a more thorough vindication of the particular entry sequence selected than had been allowed for in a previous version of this chapter.

although one that needs to be explored as an extension or alternative to the present multistage model and would require a reformulation of the solution procedure outlined. And there are a variety of other entry sequences that one might want to employ, and each will require some changes in order to solve for the SSE.

But it is important to lay emphasis on the following point. The entry sequence specified in the model is not fortuitous; it was deliberately selected because it seemed to reflect key features of firms' product development efforts and empirical regularities in industrial dynamics. First, that investment in new product innovation in all cases precedes that in imitation, which explains the order of that particular sequence. Second, that firms that have obtained an improved product design can by and large market it for a period of time before 'clones' appear in the industry. This is the rationale underlying the two separate stages in which innovating firms receive a pay-off: stage 2 corresponds to that stretch of time in which their product takes the lead in the marketplace, although they still need to compete against traditional producers, whereas in stage 3 they can face competition from product clones. And, third, that the existence of significant sunk costs for either innovative- or imitative-R&D will mean that firms that have made such investments are less likely to abandon the industry than non-innovating firms marketing a widely available product design having very narrow or no margins; this is behind the assumption that traditional firms decide their entry at each stage, and after the other firms have disbursed the requisite sunk costs. These features were prevalent in the history of DuPont, some details of which were mentioned at the very start of the chapter, but other examples may be furnished.[14]

[14] For instance, the history of IBM's product innovations in hardware and software in the early 80's, which allowed it to become the dominant supplier of microcomputers for around four years. In that time, it introduced the IBM-XT, a microcomputer with a hard disk, and later the IBM-AT, which had a much improved Central Processing Unit (CPU). But, from around 1984 onwards, PC clones that performed as well as IBM's product – and were cheaper – started to take over the market. The new technologies introduced by IBM to drive out the clones, such as the PS/2 computers, were unable to regain a dominant position in the marketplace. For a more detailed account of this history see, e.g., Hagedoorn et al (2001) or Pugh (1995).

The foregoing critique of the model's assumptions has identified diverse limitations of the analysis, but has tried to argue that these do not invalidate the general approach put forward. These stringent assumptions have been made in the interest of parsimony, and relaxing them suggests an interesting agenda for future theoretical research. For the purposes of this thesis, however, it seems more useful to retain the specified model as it stands and to give a better characterisation of the solution. That is the objective of Chapter 7, which makes a numerical analysis of the SSE. The analysis is undertaken in a setting where the numbers of innovative, traditional and imitating firms are integer-valued rather than continuous variables. One advantage of using numerical methods is that exact values for the firm-level output and pricing choices over the evolution of the industry can be computed. From these, it will be possible to construct indicators of industrial competitiveness such as price mark-ups and structural measures of concentration; as well as to calculate the aggregate profitability and consumer welfare. Last but not least, the precise influence of Knightian uncertainty on structural change, which only was touched upon in this chapter, can be explained.

Appendix. Extending the model to include "ex post participation constraints"

The multistage model developed in this chapter assumed that firms investing in innovative-R&D would actively compete in the second and third stages against any traditional and imitating firms. This allowed the notation and analysis to be simplified, as the number of innovating firms could be taken as given once investment decisions were settled, whatever the realised outcome of the innovation process. But once the uncertainty is dispelled, innovating firms may discover that their per-stage revenue is less than the fixed cost of production if they *all* compete in quantity with their improved product. It would stand to reason that some of the investing firms would prefer to exit the industry, absorbing the loss of their innovative-R&D spending. Exit would need to be such that "participating" and "exiting" firms receive exactly zero profits *ex post*. To take this eventuality into account, the model needs to include nonnegative profit conditions for innovating firms in the second and third stages. These "ex post participation constraints," also called limited liability constraints in contract theory, can be used to determine the number of innovating firms that can be supported by the market given the quality of the new product. Naturally, when considering whether or not to finance the sunk cost of innovation *ex ante*, firms will want to estimate how this contingency modifies their profits in alternative states of nature.

The purpose of this Appendix is twofold: first, to explain how the model structure, equilibrium concept, and solution procedure introduced in Section 6.2 can be generalised to accommodate an exit option for innovating firms; second, to establish conditions under which innovating firms would find participation *ex post* profitable, and those in which they would not, and investigate how exit is related to other entry and investment decisions. The analysis shows that innovating firms' ex post participation constraints (henceforth EPPCs) will always be satisfied if there is positive entry by either traditional or imitating firms. This follows

directly from the results derived in Section 6.3. If innovators' EPPCs are violated, the number of innovating firms that can make exactly zero profits (absent the sunk cost of innovation) automatically blockades entry by traditional or imitating firms. The direct consequence of these results is that the industry evolution in which EPPCs are violated can be neatly separated from the other possible evolutions we had described. Other results help to explain why including EPPCs would not invalidate the uniqueness and other results derived before. On the contrary, such an extension leads to a new outcome that has some interesting features. Before continuing, we should caution the reader that the Appendix does not have as its objective to rigorously re-work the entire model; an exhaustive description furnishing all the derivations is left for future work.

Turning first to the model structure. It is necessary to introduce an "exit" decision for innovating firms in stages 2 and 3 into the model specification set out in Section 6.2. The central assumption relates to the timing of exit, in relation to the sequence of other entry decisions. By assuming that any exit is prior to the entry of traditional and imitating firms in each stage, the basic leader-follower structure remains the same, as does the solution procedure by which the equilibrium can be found.[15] Some additional notation is needed to distinguish between the N_I firms investing in innovative-R&D, and the subset of these that actively compete in the final product market. Since the latter is determined *ex post*, it needs to be conditioned on the outcome of the innovation process and the number of innovating firms. Let $N_{I,2}(Z_i) \leq N_I$ denote the number of "participating" innovating firms in stage 2, each in possession of an improved product design of quality $Z_i = Z_L, Z_H, Z_R$. Let $N_{I,3}(Z_i) \leq N_I$ be the corresponding number of participating innovators in stage 3. Accordingly, the numbers of "exiting" innovators in stages 2 and 3 equal $N_I - N_{I,2}(Z_i)$ and $N_I - N_{I,3}(Z_i)$, respectively.

[15] Section 6.4 explains the rationale behind this sequence of entry. If the order were reversed, traditional or imitating firms would become leaders in the intra-stage sequential game. As a result, the entry decisions of these firms could force some or all of the innovators to leave the industry.

We suppose that $N_{I,2}(Z_i)$ is determined before the entry decision by traditional firms in stage 2. $N_{I,3}(Z_i)$ is determined at the very beginning of the third stage, before any investment in imitative-R&D is undertaken.

The symmetric equilibrium concept needs to be reformulated to take into consideration that some among the firms that invest in innovative-R&D may not actively compete in each stage. Making this allowance, the SSE can be defined as: the maximum number of symmetric innovating firms that can anticipate making zero profits (N_I^*) given their subjective expectations about the uncertain innovation process, their option to exit the industry at each stage, and the prospect that traditional and imitating firms will enter up to the point where they get zero profits in each stage; the maximum number of symmetric innovating firms that can participate (compete) in each stage making nonnegative profits ($N_{I,2}^*, N_{I,3}^*$), subject to the realised outcome of their investments in innovative-R&D and the prospect that traditional and imitating firms will enter up to the point where they get zero profits in each stage; and the maximum number of symmetric imitating and traditional firms that can get zero profits at each stage (N_C^*, $N_{T,1}^*$, $N_{T,2}^*$, $N_{T,3}^*$), conditional on the realised outcome of the investments in innovative-R&D and the number of innovating firms that are actually participating. In this redefinition of the SSE, the number of innovating firms is determined from firms' forecasts about the innovation process. The equilibrium number of traditional, imitating, and innovating firms competing in the industry post-innovation is assessed retrospectively, based on the equilibrium scale of investment in innovative-R&D and its actual outcome.

Recall that the solution procedure determined the SSE by combining backward recursion and zero profit and zero expected profit conditions. This method can be generalised to the problem at hand as follows. Instead of substituting the sustainable entry functions

$N_{T,3}^{MAX}$, N_{C}^{MAX}, and $N_{T,2}^{MAX}$ into the cumulative payoff of a typical innovator (see sub-step 4 on p. 235), they need to be introduced into the profit functions in stages 2 and 3, net of fixed production costs but without subtracting the sunk cost of innovation. From these one can derive the maximum number of innovating firms that can get zero profits, once each has factored in the anticipated competition from traditional and imitating firms, denoted $N_{I,2}^{MAX}(Z_i)$ for stage 2 and $N_{I,3}^{MAX}(Z_i)$ for stage 3. These sustainable entry functions can be employed to evaluate whether a given number of innovating firms can get nonnegative profits in the state $Z_i > Z_T$; this, in turn, establishes the number of innovating firms that have to exit to satisfy the EPPC. All innovators make a nonnegative net profit by competing in stage 2 if $N_I \leq N_{I,2}^{MAX}(Z_i)$; the condition for stage 3 is $N_I \leq N_{I,3}^{MAX}(Z_i)$. Conversely, if $N_I > N_{I,2}^{MAX}(Z_i)$ firms invest in innovative-R&D, then they would make a loss if they all competed in the product market. Exactly $N_{I,2}^{MAX}(Z_i)$ innovators would need to participate and $N_I - N_{I,2}^{MAX}(Z_i)$ exit for each to make zero net profits in stage 2; if $N_I > N_{I,3}^{MAX}(Z_i)$ in stage 3, then $N_{I,3}^{MAX}(Z_i)$ participate and $N_I - N_{I,3}^{MAX}(Z_i)$ exit.

The above functions summarily state the conditions under which each stage's EPPC is satisfied and those in which it is violated, and the precise partition of innovating firms into those participating and those exiting that allows each to make zero net profits in the latter case. They can be substituted into innovators profit functions in each stage. One can then continue the solution procedure outlined in the chapter to obtain the expected (net, cumulative and discounted) profits of a typical innovator, and hence the equilibrium industry-wide investment in innovative-R&D N_I^*. The SSE number of innovators explicitly takes into account that a precise number of these firms exit when the state of nature is such that their EPPCs are violated. As before, substituting N_I^* and Z_R into the sustainable entry functions

will yield N_C^*, $N_{T,2}^*$, and $N_{T,3}^*$. The same is true of the SSE number of participating firms in each stage, $N_{I,2}^*$ and $N_{I,3}^*$, the new elements of the equilibrium concept. After substitution, one obtains that: if $N_I^* \le N_{I,2}^{MAX}(Z_R)$ then $N_{I,2}^* = N_I^*$, if $N_I^* > N_{I,2}^{MAX}(Z_R)$ then $N_{I,2}^* = N_{I,2}^{MAX}(Z_R)$; if $N_I^* \le N_{I,3}^{MAX}(Z_R)$ then $N_{I,3}^* = N_I^*$, if $N_I^* > N_{I,3}^{MAX}(Z_R)$ then $N_{I,3}^* = N_{I,3}^{MAX}(Z_R)$. This completes our (condensed) explanation of the different changes needed to generalise the game-theoretic model to include EPPCs. The Appendix now turns to study the implications of this extension. Apart from the new results this exercise will provide, it may help to dispel remaining doubts about the extension.

The analysis begins by deriving the sustainable entry function for innovating firms in stage 3. $N_{I,3}^{MAX}(Z_i)$ can be determined from the net profit function of a typical innovator that has factored in the anticipated competition from traditional and imitating firms. Recall from Section 6.3 that this profit function equals:

$$\Pi_{I,3}(Z_i) =$$

$$\begin{cases} F_C & \text{if } 0 < N_I < \overline{\overline{N}}(Z_i) \text{ and } \Pi_C^{MAX}(Z_i) \ge 0 \\ 2(1-c)(Z_i - Z_T)\left[\dfrac{\sqrt{2SF_P}(4-\sigma)+S(1-c)(Z_i-Z_T)}{(4-\sigma)^2}\right] & \text{if } 0 < N_I < \overline{N}(Z_i) \text{ and } \left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \ge \overline{\overline{N}}(Z_i)\right) \\ 2S\left[\dfrac{(1-c)Z_i}{4+\sigma(N_I-1)}\right]^2 - F_P & \text{otherwise} \end{cases}$$

To find $N_{I,3}^{MAX}(Z_i)$, one must set the above expression equal to zero and solve for the number of innovators N_I. But notice that firms that invested in innovative-R&D make strictly positive profits whenever traditional or imitating firms enter in stage 3. The reason for this

quasi-rent, as we explained in Section 6.3, is that innovating firms have a competitive advantage over traditional firms from their differentiated product of quality $Z_i > Z_T$. Compared to potential imitators, they have a cost advantage, as the latter have to pay a development cost F_C to get a product of equal quality. This implies that the EPPC for innovating firms cannot be violated if: $0 < N_I < \overline{\overline{N}}(Z_i)$ and $\Pi_C^{MAX}(Z_i) \geq 0$, or $0 < N_I < \overline{N}(Z_i)$ and $\left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \geq \overline{\overline{N}}(Z_i)\right)$. Only when these conditions are not satisfied can a typical innovator's profits be zero or negative. Consequently, $N_{I,3}^{MAX}(Z_i)$ needs to be conditioned to be in the intervals of N_I and $\Pi_C^{MAX}(Z_i)$ such that traditional and imitating firms are not feasible; this calls for $\overline{N}(Z_i) \leq N_I$ and $\left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \geq \overline{\overline{N}}(Z_i)\right)$.

The exposition of the results that follow is greatly simplified by introducing a function similar to $\overline{N}(Z_i)$ and $\overline{\overline{N}}(Z_i)$, but for the innovating firms. Let $\overline{\overline{\overline{N}}}(Z_i)$ denote the number of innovators that lets each get exactly zero profits when competing in quantity with a product of quality Z_i in a given stage – conditional on there not being any traditional or imitating firms in the industry. Profits are evaluated net of the fixed production cost but before subtracting the sunk innovation cost. This will be the number of innovating firms that solves the following equation:

$$\pi^{CN}(Z_i|0, N_I, 0, Z_i) - F_P = 2S\left[\frac{(1-c)Z_i}{4+\sigma(N_I-1)}\right]^2 - F_P = 0$$

The solution can be shown to be:

$$\overline{\overline{\overline{N}}}(Z_i) = \sqrt{\frac{2S}{F_P}}\frac{(1-c)Z_i}{\sigma} - \frac{4-\sigma}{\sigma}$$

$$= \frac{4-\sigma}{\sigma} \frac{\sqrt{2S}(1-c)Z_i - \sqrt{F_P}(4-\sigma)}{\sqrt{F_P}(4-\sigma)}$$

Using this notation, it is clear that $N_{I,3}^{MAX}(Z_i) = \overline{\overline{\overline{N}}}(Z_i)$ *if* $\overline{N}(Z_i) \leq N_I$ and $\left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \geq \overline{\overline{\overline{N}}}(Z_i)\right)$. In the other two intervals there is no N_I such that innovators' net profits equal zero in stage 3 and therefore $N_{I,3}^{MAX}(Z_i)$ cannot be defined: in these intervals the innovating firms can all make a positive profit by participating. A comparison of $\overline{\overline{\overline{N}}}(Z_i)$ with $\overline{N}(Z_i)$ and with $\overline{\overline{N}}(Z_i)$ shows that this result does not give rise to a contradiction. By comparing $\overline{\overline{\overline{N}}}(Z_i)$ with $\overline{N}(Z_i)$, one can immediately establish that $\overline{\overline{\overline{N}}}(Z_i) > \overline{N}(Z_i)$ for any $Z_i > Z_T$:

$$\overline{\overline{\overline{N}}}(Z_i) = \frac{4-\sigma}{\sigma} \frac{\sqrt{2S}(1-c)Z_i - \sqrt{F_P}(4-\sigma)}{\sqrt{F_P}(4-\sigma)} > \frac{4-\sigma}{\sigma} \frac{\sqrt{2S}(1-c)Z_T - \sqrt{F_P}(4-\sigma)}{\sqrt{2S}(1-c)(Z_i - Z_T) + \sqrt{F_P}(4-\sigma)} = \overline{N}(Z_i)$$

$$=$$

It is apparent that the numerator of $\overline{N}(Z_i)$ is greater than that of $\overline{N}(Z_i)$, and its denominator is smaller, for any $Z_i > Z_T$. A comparison of $\overline{\overline{\overline{N}}}(Z_i)$ and $\overline{\overline{N}}(Z_i)$ directly shows that $\overline{\overline{\overline{N}}}(Z_i) > \overline{\overline{N}}(Z_i)$ for any $F_C > 0$:

$$\overline{\overline{\overline{N}}}(Z_i) = \frac{4-\sigma}{\sigma} \frac{\sqrt{2S}(1-c)Z_i - \sqrt{F_P}(4-\sigma)}{\sqrt{F_P}(4-\sigma)} > \frac{(4-\sigma)}{\sigma} \frac{\sqrt{2S}(1-c)Z_i - \sqrt{F_P + F_C}(4-\sigma)}{\sqrt{F_P + F_C}(4-\sigma)} = \overline{\overline{N}}(Z_i)$$

Taken together, these results imply that when $0 < N_I \leq \overline{\overline{\overline{N}}}(Z_i)$ firms invest in innovative-R&D, they will all actively compete in stage 3 making a nonnegative profit. Entry by traditional and imitating firms may or may not be supported in the SSE, this depends on the exact number of firms investing in innovative-R&D and the realised outcome of their

research efforts. The innovators' EPPC will be violated in the third stage if $N_I > \overline{\overline{N}}(Z_i)$, in which case $\overline{\overline{N}}(Z_i)$ innovators participate and $N_I - \overline{\overline{N}}(Z_i)$ exit, and each gets zero profits. Since $\overline{\overline{N}}(Z_i) > \overline{N}(Z_i)$ for any $Z_i > Z_T$, and $\overline{\overline{N}}(Z_i) > \overline{N}(Z_i)$ for any $F_C > 0$, the number of participating innovating firms when the EPPC is violated $\overline{\overline{N}}(Z_i)$ will automatically blockade entry by traditional and imitating firms in stage 3. The basis for this result is the sequence of entry specified, which has innovating firms in the lead position, and the quality and cost advantages of these firms vis-à-vis those that would produce the standard product and potential imitators.

In relation to the second stage, the expression for a typical innovator's profit, once this factors in the anticipated entry by traditional firms, was found to be:

$$\Pi_{I,2}(Z_i) =$$

$$\begin{cases} 2(1-c)(Z_i - Z_T)\left[\dfrac{\sqrt{2SF_P}(4-\sigma)+S(1-c)(Z_i - Z_T)}{(4-\sigma)^2}\right] - F_I & \text{if } 0 < N_I < \overline{N}(Z_i) \\ 2S\left[\dfrac{(1-c)Z_i}{4+\sigma(N_I - 1)}\right]^2 - F_I - F_P & \text{if } N_I \geq \overline{N}(Z_i) \end{cases}$$

This expression subtracts the sunk cost of innovation F_I. However, when we come to evaluate innovators' EPPC, this investment has already been disbursed and cannot be recuperated. Thus, the important question is whether profits net of fixed production costs are nonnegative or not. Without deducting the sunk cost of innovation, it is clear that $0 < N_I < \overline{N}(Z_i)$ innovating firms make a positive profit, corresponding to their quasi-rent in conditions where there is positive entry by traditional firms. Innovating firms may make a loss if they all participate when $N_I \geq \overline{N}(Z_i)$. Notice that, absent the sunk cost of innovation,

the expression for profits in this interval is identical to that in the third stage. In consequence, it is easy to verify that $N_{I,2}^{MAX}(Z_i) = \overline{\overline{N}}(Z_i)$ *if* $\overline{N}(Z_i) \leq N_I$, and that it is not defined if $0 < N_I < \overline{N}(Z_i)$. Again, the implication is that $0 < N_I \leq \overline{\overline{N}}(Z_i)$ firms investing in innovative-R&D can all compete in stage 2 without contravening their EPPC. If the number of innovators is $N_I > \overline{\overline{N}}(Z_i)$, then $\overline{\overline{N}}(Z_i)$ participate and $N_I - \overline{\overline{N}}(Z_i)$ exit to precisely obtain zero profits *ex post* with a product of quality $Z_i > Z_T$. Traditional and imitating firms would be automatically blockaded in any SSE having positive exit by innovators.

To sum up the results obtained thus far. When $0 < N_I \leq \overline{\overline{N}}(Z_i)$ firms invest in innovative-R&D, they can all actively compete in stages 2 and 3 making a nonnegative per-stage profit that is offset against the sunk cost of innovation and can lead to an overall profit. The innovators' EPPC will be violated if $N_I > \overline{\overline{N}}(Z_i)$, in which case $\overline{\overline{N}}(Z_i)$ innovators participate and $N_I - \overline{\overline{N}}(Z_i)$ exit in both stages. This is the extent of exit that makes innovating firms indifferent between actively competing and exiting from the industry. By definition, participating firms obtain exactly zero profits net of fixed production costs from competing in each stage, and would make a cumulative loss of $-F_I$. The exiting firms would lose $-F_I$ as well. These simple results imply that the number of innovating firms satisfying the EPPC is the same in each stage, whether there is exit or not: either they all compete in both stages, or the same number will exit. Additionally, exit by innovating firms is never so great as to create the conditions for subsequent entry by traditional or imitating firms: non-innovators will find that they are automatically blockaded whenever the market does not support all innovating firms.

The feasible *ex post* participation of innovating firms can be substituted into the profit functions of a typical innovator in the second and third stages. The expressions thereby obtained explicitly account for any exit to satisfy the EPPCs and the anticipated competition from firms marketing standard or cloned products. In stage 3, this profit function equals:

$$\Pi_{I,3}(Z_i) =$$

$$\begin{cases} F_C & \text{if } N_I < \overline{\overline{N}}(Z_i) \text{ and } \Pi_C^{MAX}(Z_i) \geq 0 \\ 2(1-c)(Z_i - Z_T)\left[\dfrac{\sqrt{2SF_P}(4-\sigma)+S(1-c)(Z_i - Z_T)}{(4-\sigma)^2}\right] & \text{if } N_I < \overline{N}(Z_i) \text{ and } \left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \geq \overline{\overline{N}}(Z_i)\right) \\ 2S\left[\dfrac{(1-c)Z_i}{4+\sigma(N_I-1)}\right]^2 - F_P & \text{if } \overline{N}(Z_i) \leq N_I < \overline{\overline{N}}(Z_i) \text{ and } \left(\Pi_C^{MAX}(Z_i) < 0 \text{ or } N_I \geq \overline{\overline{N}}(Z_i)\right) \\ 0 & \text{if } N_I \geq \overline{\overline{N}}(Z_i) \end{cases}$$

In stage 2, the substitution yields the following function:

$$\Pi_{I,2}(Z_i) =$$

$$\begin{cases} 2(1-c)(Z_i - Z_T)\left[\dfrac{\sqrt{2SF_P}(4-\sigma)+S(1-c)(Z_i - Z_T)}{(4-\sigma)^2}\right] - F_I & \text{if } 0 < N_I < \overline{N}(Z_i) \\ 2S\left[\dfrac{(1-c)Z_i}{4+\sigma(N_I-1)}\right]^2 - F_I - F_P & \text{if } \overline{N}(Z_i) \leq N_I < \overline{\overline{N}}(Z_i) \\ -F_I & \text{if } N_I \geq \overline{\overline{N}}(Z_i) \end{cases}$$

In contrast to the profit functions derived under the assumption that all innovating firms compete *ex post*, these functions are bound from below. The lower bound equals 0 in the third stage and $-F_I$ in the second, the amount spent by each firm on irreversible innovative-R&D. This is the maximum liability faced by innovating firms having an exit option, however poor their research efforts turn out to be. Thus, we find that the profit curve in each stage would be characterised by: a flat segment in which there is entry by traditional or imitating firms; a declining segment when the number of innovating firms automatically blockades traditional and imitating firms, or the latter find that entry is too costly $\Pi_C^{MAX}(Z_i)<0$, but in which all innovators can profitably participate; and lastly, a flat segment corresponding to "excess entry" by innovating firms, in which a precise number of these firms must exit so that each makes zero profit (not accounting for the sunk innovation cost), and no entry by traditional or imitating firms is feasible.

The final and, in theoretical terms, most important question is whether the generalised SSE is unique for $N_I^* > 0$. Recall that the uniqueness of the SSE rested on the properties of continuity and monotonicity of the profit and expected profit functions for innovating, traditional and imitating firms. The foregoing analysis indicates that a typical innovator's profit functions in stages 2 and 3 continue to be continuous and monotonically decreasing after making allowances for exit when the EPPCs are contravened. Because these properties are preserved under summation, the expect profit function will be continuous and monotonically declining, which implies that the SSE number of innovators will be unique. Clearly, N_I^* can never be so great that innovators expect to exit in both states of nature considered *ex ante*, since their expected profit would then be $-F_I$. They can anticipate that some among them will need to exit in the state where they obtain a minor innovation, but this needs to be compensated by a positive gain when it turns out they can market a major innovation. The SSE number of participating innovators will be the same in stages 2 and 3,

$N^*_{I,2} = N^*_{I,3}$, whether all or just some firms compete in equilibrium. The uniqueness of $\overline{\overline{N}}(Z_i)$ for any $Z_i > Z_T$ including Z_R and that of N^*_I are sufficient for $N^*_{I,2} = N^*_{I,3}$ to be unique.

The number of traditional firms in the stage 1 SSE ($N^*_{T,1}$) is completely independent of subsequent product development and would still be unique. Concerning the uniqueness of traditional ($N^*_{T,2}$, $N^*_{T,3}$) and imitating firms (N^*_C) in stages 2 and 3, observe that the introduction of EPPCs has no effect whatsoever on the profit functions of traditional and imitating firms, for two reasons. First, because the innovating firms decide whether, and how much, exit is necessary before entry by traditional and imitating firms is determined in each stage. The leader-follower structure is therefore unchanged. Second, the analysis has shown that: the EPPCs cannot be violated if there is positive entry by traditional or imitating firms; and the case in which the EPPCs are violated cannot support entry by traditional or imitating firms. This implies that *ex post* exit by innovating firms can have no effect on subsequent investment in imitative-R&D or competition with the standard product. The profit functions of traditional and imitating firms are related to the number of innovators exactly as in the model without EPPCs, and therefore the uniqueness of N^*_I guarantees that of N^*_C, $N^*_{T,2}$, and $N^*_{T,3}$. The industry evolution in which the exit option is 'exercised' by innovating firms after learning the outcome of their R&D efforts is clearly separated from those that had been identified in the chapter.

There are two main conclusions to be drawn from the analysis in this Appendix. The first is that the results we derived in the main body of the chapter are valid if the multistage model is extended to include EPPCs. In particular, the uniqueness of an SSE with positive investments in innovative-R&D would extend to the case in which EPPCs need to be satisfied – once the equilibrium concept is generalised to give innovators an exit option in stages 2 and 3. The second conclusion is that a model with EPPCs can generate an industry evolution that

is fundamentally different to those that had been identified in Section 6.3. The evolution in which the EPPCs are violated can be characterised as an investment boom-bust cycle that is driven by forecast errors concerning the outcomes of uncertain innovative-R&D. Specifically, the evolution would have some innovators permanently exit the industry, the standard product would be forced out from stage 2, and no imitative-R&D would be financed. Each of the innovating firms would absorb a loss equal to their sunk cost of innovation, the maximum amount they are liable for if they can choose at each stage to not compete in the final product market. This adds to the richness of the results of this theoretical approach, since it reveals conditions under which firms investing in innovation can force traditional firms from the industry and yet become victims of their "creative destruction."

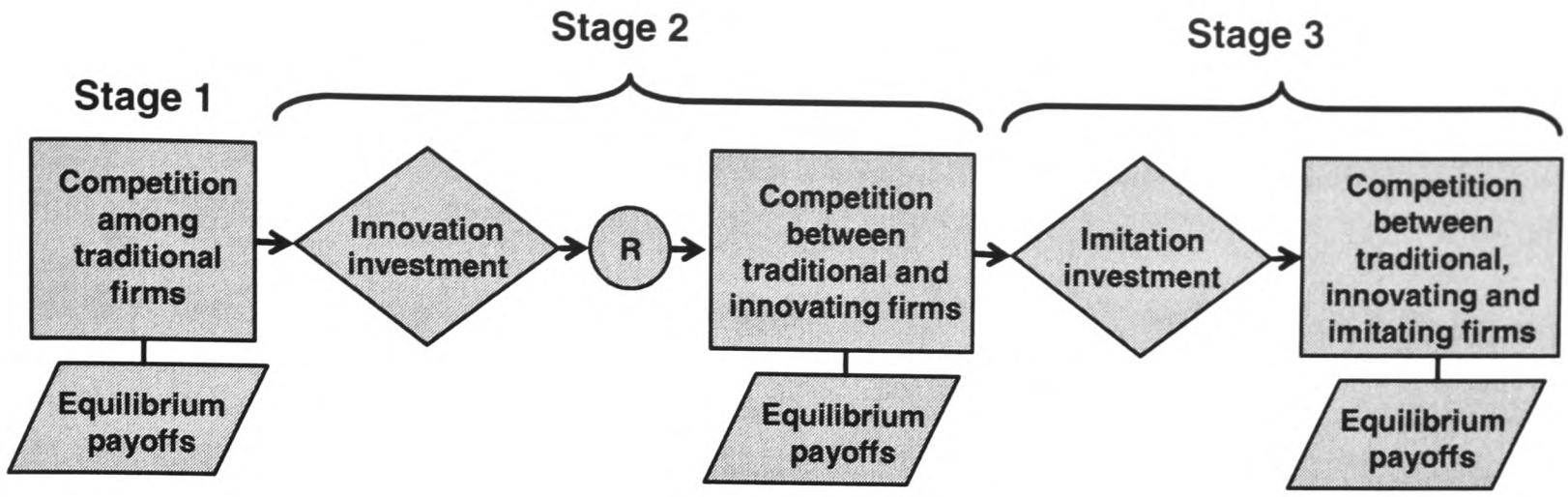
Chapter 7

Dynamics of industry structure and market competition under uncertainty: A numerical analysis

7.1 Introduction

With the purpose of investigating the evolving industrial structure arising from successive waves of innovation and imitation, Chapter 6 introduced a game theoretic approach featuring two methodological novelties: the explicit analysis of fundamental uncertainty surrounding technological advances, in which firms' profit expectations and investment decisions depended to a degree on subjective evaluation of opportunities that could not be reduced to empirical risk assessments; and the endogenous determination of the market structure and R&D expenditures in a dynamic setting where the details of the sequence of "entry decisions" mattered. In the multistage model to which this approach led, R&D investments in product development were made sequentially, starting with a stage in which firms decided whether or not to fund new product research, the outcome of which was fundamentally uncertain. The decision process was represented by a subjective state-contingent pay-off matrix, in which the pay-off to alternative strategies depended upon the subsequent evolution of firm behaviours once the fundamental uncertainty was dispelled. Other firms waited until the uncertainty dissipated before investing in costly but certain imitation of the improved product design. The investments were made in a vertically and horizontally differentiated market, in which innovators and imitators faced effective competition from existing or potential non-innovating firms manufacturing a conventional product. The general structure of the model is depicted in Figure 1; a detailed description is contained in the second section of Chapter 6.

Figure 7.1 The sequential structure of the model



After setting out the general structure and specific assumptions of the sequential model, Chapter 6 put forward a novel equilibrium concept and a recursive procedure by which this equilibrium could be found. The "symmetric sustainable equilibrium" was defined as: the maximum number of symmetric innovating firms expecting to make zero profits given their subjective probability estimates about the uncertain innovation process and the prospect that traditional and imitating firms enter up to the point where they get zero profits in each stage; and the maximum number of symmetric imitating and traditional firms that can get zero profits at each stage, conditional on prior investments in innovative-R&D and the realised outcome of these investments. A formal analysis proved that, if a symmetric sustainable equilibrium with positive investment in innovation existed, it would be unique. The analysis also established the range of industry evolutions that could be supported and the conditions under which each evolution would arise.

In this chapter, these formal results are supplemented by an exact numerical solution of the model in a setting where the numbers of innovating, imitating, and traditional firms are integer-valued. One objective of this numerical analysis is to examine the tendencies in product development, market structure, pricing, firm growth, profitability, and consumer welfare over the evolution of the industry. Specifically, we mean to address the following questions: What effect does innovative-R&D have on structural and performance-based

measures of market power? Is the influence of innovation on market structure undone or reinforced by investments in imitative-R&D? Do firms grow on average or do they become smaller? Is the "product lead time" sufficient for innovating firms to break even, or not? What are the trends in industry-level indicators of sales and profitability? Is it innovation or imitation that drives consumer and social welfare increases? Connected to this, the second objective is to evaluate the strength of the reciprocal relations between the firms' disposition towards uncertainty, the concrete outcomes resulting from innovation process, and the feasible industry evolutions. That is, in what respects are the answers to each of the previous questions contingent on an industry's *ex ante* attitude towards uncertainty and innovators' success *ex post*?

To answer this list of research questions, we compare two alternative industry evolutions: one in which all the innovating firms obtain a minor innovation, and another where they all get a major innovation. In other words, innovators learn that the realised outcome of their investments coincides with one of the two states they had originally forecasted after the uncertainty is dispelled. Although this assumption is not necessary (the equilibrium can just as easily be computed in situations where innovators find themselves with a product innovation different in quality to what they had foreseen), there is no great loss in generality from making it. This is because there is only one outcome in which the equilibrium number of innovating firms – i.e., that derived from a zero expected profit condition – exactly allows each to make zero profits *ex post*; and notice that this exceptional outcome cannot be one of the two that the firms were expecting at the outset. For all other outcomes, there will have been too little or too much spending on innovative-R&D, in the sense that the firms funding this mode of research will receive positive or negative profits. Given this feature of the equilibrium, one can get a reasonable indication of the solution's characteristic properties by comparing just these two industry evolutions.

The rest of the chapter is divided into three sections. Section 7.2 presents and discusses the main results from the numerical analysis. The simulations reveal a fall in competitiveness directly after innovation (in stage 2), but suggest that entry by imitating firms (in stage 3) can revert this tendency. Concentration may follow an inverted U-shaped path when innovative-R&D has produced a major innovation. To the extent that clones of the product become available, technological leaders experience a decline in market shares and there is "convergence of returns" at the firm- and industry-level. These results seem to be consistent with many empirical regularities pertaining to entry, exit, market structure, and technological change identified in the development of new industries.[1] Further, our analysis suggests that success in innovation benefits consumers on two counts: there is an initial increase in welfare stemming from purchases of the enhanced quality good from innovators, and a further gain as a "swarm of imitators" introduce duplicates that drive down prices. Conversely, a small improvement in quality brings marginal gains to consumers, even when product clones are marketed, as consumers face a progressively more concentrated market structure and higher prices in exchange for a small increase in product quality. Section 7.3 examines the effects of the industry's disposition towards uncertainty, their "optimism or "pessimism," on the alternative industry evolutions. The last section offers some concluding remarks about the merits of this framework in the light of related approaches.

7.2 Numerical solution

7.2.1 Methodology

To solve the model numerically, the equations for net profits evaluated at the Cournot-Nash equilibrium were introduced into an Excel spreadsheet and the solution procedure we formalised was carried out step by step to get the sustainable entry sequences for specific

[1] As described in, e.g., Klepper (1996), Klepper and Graddy (1990) or Caves (1998).

parameter values (listed below). Unlike the formal analysis undertaken in Chapter 6, in which the numbers of innovating, imitating and traditional firms were continuous variables for reasons of tractability, numerical methods are well adapted to the calculation of the more realistic integer-valued solution to the problem. All that is required is to switch from zero profit (equality) conditions to nonnegative profit (weak inequality) conditions throughout the solution process. Excel's Solver was employed to determine the "symmetric sustainable equilibrium"; this solves constrained optimisation problems with non-linear objective functions through a Generalized Reduced Gradient method. In finding an answer to the problem in terms of integer values for the number of firms, the exact solutions derived by the Excel Solver for each zero profit condition were rounded down at each step of the process.

Bearing in mind that the uniqueness of the "symmetric sustainable equilibrium" for any positive investment in innovation has been established in the continuous case, this numerical technique seems powerful enough. But one needs to be aware that, in the limiting cases where the flat segment of the profit functions derived in Chapter 6 approximately coincides with the zero axis, one may get an equilibrium number of firms in the middle of the segment rather than at the rightmost point that was the maximum with continuous variables. This is because the profit curves will exhibit small variations around the theoretically flat segment when integers are used; Figure 7.2 depicts a typical innovator's profit curves, which shows that the size of the variations is very small indeed. At any rate, when running each simulation, care was taken to avoid limiting cases where these variations generate a dynamic sequence that would not have been supported when working with continuous variables.

The results presented in this section are derived from simulations with a particular set of parameter values, which is labelled the 'benchmark' case. Although the trends revealed by these simulations are conditioned by the parameter values we will assume, the generality of these patterns should become clear when the results of the comparative equilibrium dynamics

analysis with respect to various parameters are introduced in the course of the discussion, in Section 7.3 and also in Chapter 8. The benchmark case assumes intermediate values for the degree of horizontal differentiation $\sigma = 1$, the discount rate $\delta = 0.8$, the marginal cost $c = 0.5$, and the estimated probabilities of a minor and a major innovation $1 - p^* = p^* = 0.5$. The market is made up of 100,000 representative consumers with the same quality-augmented quadratic utility function described previously. A minor innovation has a product quality that is 25 % above the standard product, while the major innovation is a 100 % improvement ($Z_T = 1$, $Z_L = 1.25$, $Z_H = 2$). The sunk cost of innovation is five times that of imitation, and twenty five times the fixed cost of production ($F_I = 2500$, $F_C = 500$, $F_P = 100$); to get a sense of the magnitude of these costs, note that the equilibrium per stage revenue of a traditional manufacturer before innovation turns out to be around 1,250. Appendix 7.1 lists the parameter values assumed in the simulations discussed in this chapter.

7.2.2 Product development and market structure

The numbers of companies innovating, imitating and competing with the standard product in the simulations with the benchmark parameters are shown in Table 7.1, which sets side by side the sustainable industrial structures for the two alternative outcomes of the innovation process. The first point to note is that the number of traditional firms at stage 1 is high, indicating that the industry structure is quite competitive prior to the innovations' appearance. Second, the number of innovating firms is positive – four firms disburse the fixed cost F_I. This number evidently is the same whatever the outcome of the investment, because the latter decision must be taken under uncertainty and in advance. But the nature of the outcome is vital in determining the number of traditional manufacturers that will remain in the market (seven companies for a minor innovation vs. zero for a major one); and also the number of

firms that will clone the product (eight for a minor innovation, and almost twice that when the product is a major improvement over the standard product). The contrast between the two dynamic sequences is readily explained: competition from four companies producing a minor innovation is not as severe as that from four companies with a seriously better product, nor is a slightly better version of the standard product as attractive to copy when imitation is costly.[2]

Table 7.1 Number of firms in the industry

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Traditional firms	19	7	0	19	0	0
Innovators		4	4		4	4
Imitators			4			11
Total	19	11	8	19	4	15

Since the model does not keep track of the identity of the firms at each stage, a variety of specific "histories" can be composed in a fashion that is consistent with the evolution of the industry equilibrium. One possible sequence of events is for the 4 firms investing in own-product R&D to be a subset of the 19 traditional manufacturers that coexist in the market in the pre-innovation equilibrium. Once the uncertainty is resolved, these firms switch to the new product. The remaining 15 firms then take production and development decisions depending on the type of product they must compete against: when it is marginally superior to the established one, seven firms prefer to stay in the market with the standard product and exit later on, 4 firms imitate the new good and thus assure their long-term survival, and the other 4 will close their doors permanently. When there is a great improvement in the product

[2] In this respect, the results agree with Nathan Rosenberg's (1996, p. 347) comments on major innovations: "Major innovations, such as the steam engine, once they have been established, have the effect of inducing further innovations and investments over a wide frontier. Indeed, the ability to induce such further innovations and investments is a reasonably good definition of what constitutes a major innovation. It is a useful way of distinguishing between technological advances that are merely invested with great novelty from advances that have the potential for a major economic impact. But this also highlights the difficulties in foreseeing the eventual impact, since they will depend on the size and direction of these future complementary innovations and associated investments."

quality design, then 11 of the incumbent firms that initially did not innovate will create clones of this good, and the rest will leave the industry. Of course, a different account of the industrial dynamics could be made, in which the businesses that are financing innovation are mostly or completely entrants rather than incumbents.

As the total number of firms actively producing in a market is a commonly accepted (inverse) measure of the market's concentration, the evolution of market structure can be examined directly from Table 7.1. Alternatively, one can resort to the reputedly more robust Herfindahl-Hirschman index of concentration (in Table 7.2), which takes into account the relative size and distribution of the firms. Both statistics indicate that product innovation leads to a greater concentration of the market. The reason for the decline in industrial competitiveness is that traditional products compete at a disadvantage with the new designs and as their share in the profits of the oligopolic industry drops, the returns of traditional manufacturers are no longer enough to finance the fixed cost of production faced by each. Naturally, the standard product is at a greater disadvantage when a major innovation has been discovered, which results in the exit of the traditional vendors, and a more concentrated industry structure.

Table 7.2 Herfindahl-Hirschman index

	Minor Innovation			Major Innovation		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
H-H index	526	1,260	1,250	526	2,500	667

But the solution presented in Tables 7.1 and 7.2 reveals a countervailing force that can undo the process of concentration sparked by a product innovation. When a better product has been developed, the entry of myriad clone-producing firms – Schumpeter's "swarm of imitators" – can be a powerful agent of deconcentration; a major innovation multiplies the return from imitating (which, with a fixed cost of imitation) prompts more firms to undertake the

financing of imitative activities in a zero-profit equilibrium, intensifying competition in the third stage. In other words, a major innovation is very successful in dislodging the established products but will also be responsible for setting off a burst of investment in imitative-R&D, limiting the market shares of the technological leaders that invested in innovative-R&D.

Table 7.3 Price mark-ups[3]

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Standard product	9 %	10 %	---	9 %	---	---
Improved product		21 %	18 %		29 %	11 %
Average	9 %	18 %	18 %	9 %	29 %	11 %

Structural measures of market power such as the Herfindahl index seek to quantify the degree of market concentration, but performance-based measures can be more informative about the effective market power held by firms. One such measure is the equilibrium price mark-up for each kind of product, which indicates the deviation of unit price from the marginal cost. Table 7.3 displays the percentage price mark-up for the standard product and the improved one at each stage, for both minor and major innovations. As was the case with the number of firms, it seems that more successful outcomes from the innovation process encourage a transiently higher degree of market power; this effect is largely dissipated once the cloned products become available in the market during the third stage. When the new product quality design is a minor improvement, consumers face permanently higher price mark-ups. The lasting price hike in this case masks a more intricate evolution in pricing characterised by continued sales of the standard good at a price almost equal to the pre-innovation stage; the dispersion in prices ends when the firms copy the new good and drive the standard product from the market.

[3] The mark-up is calculated as $100*(P-c)/c$, where P is the equilibrium price of the standard or improved product, and c is the common marginal cost. The weight used to calculate the average mark-up is the equilibrium output of each kind of firm.

The joint examination of these trends suggests an interesting pattern in the observed evolution of market structure, output and pricing. In the context of technological change, a slow shakeout of traditional manufacturers combined with little imitation and an increasingly concentrated market structure could be a *sign of failure* in the firms' own-product R&D efforts. Relatively unsuccessful attempts to innovate would also result in permanently higher price mark-ups. By contrast, success would trigger an intense shakeout of the established products and transitorily the market would become very concentrated, but a major new product would also encourage the emergence of many clones, and the market would tend to become more competitive as time elapsed. If the innovators are successful, the result is a more profound concentration in the short run (that is, during the span of time in which innovating firms are marketing the improved commodity without competition from imitators); however, when other firms find out the value of the new product and act by imitating it and selling it in the final product market, the industry can potentially become more competitive than before innovation took place.

7.2.3 Firm-level trends in output and profitability

A different facet of the emerging industrial organization is revealed by an inspection of firms' size, measured in Table 7.4 by the quantities sold (but note that the same trends emerge when revenue is used as a proxy of size). The numerical solution suggests that innovative companies will be larger than traditional producers competing in the same market, and that this difference in size is positively related to the novelty of their product design. Because of the quadratic form of the utility functions, the percentage increases in average output are significantly greater than the percentage increases in product quality; in this benchmark case, a 25% increase in the product quality design provokes a two-fold increase in the average output of the firms in the 2nd stage equilibrium, whereas a 100% increase in product quality

leads to firms more than 12 times larger than before. The intense rivalry that can be expected in the case of a major innovation, in which many imitators begin to manufacture a product of the same quality as the original innovators, can reverse this tendency and lead to smaller sized firms. For a minor innovation the opposite is true, because the imitators drive out the small firms that are producing the standard good, and so the firms competing in stage 3 would tend to grow in size (on average).

Table 7.4 Output levels

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Traditional firm	2,273	2,381	---	2,273	---	---
Innovator		8,185	7,102		28,571	11,111
Imitator			7,102			11,111
Average	2,273	4,492	7,102	2,273	28,571	11,111

The types of products against which firms must compete, and the restraint put on the equilibrium price and quantity by this competition, are not as important to the firms' assessment of their success as are the profits they finally will be able to extract. Table 7.5 measures individual firm's profit margins (net of marginal and fixed costs), and the average margin calculated from industry-wide profits and revenue. By construction, the profit margins of any traditional manufacturer active in the industry and all imitators should be very close to zero, and any deviation is a result of restricting the simulations to integer values for the number of firms – the table simply confirms this for the benchmark parameters. The profit margin of a typical innovator depends on the outcome of innovation. When competing with a marginally improved product, it has sharp losses at the outset but makes a profit later on; with a major innovation, the profit margin is positive in both stages, but there is a decline

corresponding to the influx of imitators into the market in stage 2. [4] As could be expected, the figure for the overall profits (not shown in the table), is negative for a minor innovation and positive for a major one.

Table 7.5 Firm-level profit/loss margins[5]

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Traditional firm	0 %	1 %	---	0 %	---	---
Innovator		– 35 %	13 %		8 %	8 %
Imitator			1 %			0 %
Average	0 %	– 24 %	7 %	0 %	8 %	2 %

It is puzzling that the fluctuations in "financial health," as measured by profit margins, do not match the movements in the indicators of market power. For example, with a minor innovation, an innovator's profit margins increase over time even as the price mark-up falls. To take another example, after the clones enter the market, the profit margin of a typical innovator and the industry-wide margin turn out to be higher when own-product R&D falls short of firm's expectations: the innovator would register a margin of 13% with the inferior product as against 8% with a radical improvement, while the industry measure would be 7% vs. 2% respectively. This apparent anomaly can be explained by recalling that the fixed cost of innovation F_I is paid up-front, which will tend to intensify the losses *accounted for* early on with a minor innovation and depress the level of profits with a major innovation. By the same token, innovating firms get a profit in stage 3 because they can manufacture the improved good without paying further development costs.[6] This system of accounting for

[4] Other simulations show that innovators could experience negative profit margins in stage 2, but of course the cumulative discounted profit would still be positive.

[5] The profit margin results from dividing a firm's profits net of all fixed costs by the revenue that firm earns.

[6] As we showed in Chapter 6, each gets a quasi-rent equal to the imitation cost when clones of the product enter the market, and to their Cournot-Nash profit given the number of traditional manufacturers when no imitation is forthcoming. Since innovating firms earn less revenue with a minor innovation, the profit rate corresponding to this quasi-rent is higher in the low state of nature.

costs and revenue would result in a very imperfect correlation between the standard measures of market power and profit margins. The results suggest that caution should be exercised when using cross-section analysis (based on data at one point in time) to study the reciprocal effects between innovating firms' success, their financial performance and market power.

7.2.4 Industry-level outcomes

Having discussed the firm-level equilibrium decisions and outcomes, it is appropriate to investigate what these give rise to in aggregate terms, both for the industry taken as a whole as well as for consumers. It is no surprise that the industry-wide reckoning of profits in Table 7.6 shows the same trends as profit margins for an average firm: profits are close to zero before innovation; when the outcome of own-product research and development is mediocre there is a large shortfall in stage 2 followed by a small gain in stage 3; whereas, when research is very successful profits are positive but declining across stages 2 and 3. The aggregate profits are almost equal in stage 3, regardless of the outcome of innovation. This would be true of any solution exhibiting some imitation of the minor product improvement because, as we observed before, the rents gained by each innovator in stage 3 (approximately) equal the cost of imitation.[7] The progression in sales volume exhibits marked differences depending on the kind of innovation in question: in the simulations, a minor innovation leads to an increase in sales of the order of 20% from each stage to the next, whereas a major innovation results in a three-fold increase in sales from stage 1 to 2, followed by a another hike of around 30% when clones enter the market. Notice that the ups and downs in average sales experienced by the firms in the industry are not replicated at the industry-level, which is characterised by a continuous rise in sales volume.

[7] For the benchmark values, there are 4 innovators, each getting $F_C = 500$; the deviations of aggregate profits from the 2000 mark are a result of using integers in the calculations.

Table 7.6 Aggregate measures of profitability and welfare

	Minor Innovation			Major Innovation		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Innovators' profits		– 6,970	2,183		5,927	2,069
Industrial profits	63	– 6,877	2,365	63	5,927	2,259
Consumer welfare	9,814	10,240	11,622	9,814	20,408	37,037
Social welfare	9,877	3,363	13,987	9,877	26,335	39,296
Sales volume[8]	23,554	28,926	33,574	23,554	73,469	92,593

To gauge how the changes in the type of product offered in the market and the terms set upon its sale affect consumers, it is best to look directly at the value of consumer welfare in each stage. When the new product is markedly better, Table 7.6 reveals a jump in consumer welfare as consumers buy a greater volume of a much-improved product, even if these are sold at a substantially higher price. But, for the parameter values analysed, the larger jump in welfare occurs when the market is overrun by the "swarm of imitators"; thanks to the competition from these firms, the end price of the commodity will be just slightly higher than what they paid for the original, lower-quality good. To put this in another way, the increase in consumer welfare may depend as much on the existence of an improved product, as on the possibility that other firms will introduce duplicates that drive down prices in the market. Since superior products encourage more firms to disburse the cost of imitation, it is apparent that success in the innovation process benefits consumers on two different counts. It would seem that to society, "imitation, as well as nature, is the source of particular virtues."

The figures in the table indicate that a small improvement in quality brings only marginal welfare gains to consumers, even when product clones become available later on. The reason is that consumers face a progressively more concentrated market structure and permanently higher prices in exchange for a relatively small increase in product quality. When the loss suffered by the industry is factored in, the picture is bleak: social welfare just

[8] The sales volume is the total revenue earned by the firms in the industry.

after innovation is much lower than in the stage before innovation, and it is clear that the small increase in social welfare estimated for stage 3 will not compensate this deficit. Whatever the social discount rate, in this benchmark case the net effect to society is negative: even if social welfare in each stage is added with no discount at all, a stagnant, unchanging industry generates 29,631 in welfare (3 x 9,877), against 27,227 for an innovating industry in a state of flux, *but* one characterised by a relative failure on the part of researchers to reach firms *ex ante* expectations. The next section contends that the scale and distribution of the social loss is intimately connected to the way in which firms deal with the uncertainty in the innovation process. Just as a rush of imitative activity following a successful innovation can boost social welfare gains, a burst of innovative activity motivated by 'lofty expectations' can cause social losses to spiral downwards following an ineffectual attempt to innovate.

7.3 Fundamental uncertainty and structural change: A closer inspection

This section examines in more detail the dynamic interactions between firms' disposition towards the element of uncertainty surrounding innovation, the viability of the firms' alternative research strategies (innovative or imitative), and the divergent evolutions followed by the industry once uncertainty has been dispelled. To this end, recall that we modelled fundamental uncertainty by making the following assumptions. The state of knowledge only lets firms estimate that p, the 'true' probability of getting a major innovation, lies in the range $\left[\underline{p}, \overline{p}\right]$, and the probability of a minor innovation, $1-p$ lies in $\left[1-\overline{p}, 1-\underline{p}\right]$. On the basis of subjective factors, firms select a point-estimate in order to assess the expected profits from innovating: a firm that is 'optimistic' believes the correct probabilities are $p^* = \overline{p}$ and $1-p^* = 1-\overline{p}$, and a 'pessimistic' firm believes they are $p^* = \underline{p}$ and $1-p^* = 1-\underline{p}$. Lastly, the belief structure is symmetric, allowing the analysis to concentrate on two settings, in

which all firms are optimistic or all are pessimistic. For the sake of exposition, we shall consider a concrete example, where p and $1-p$ lie in the interval $[.4,.6]$. This implies that firms in a state of 'systemic optimism' believe that $p^* = \overline{p}$ =0.6, and $1-p^* = 1-\overline{p}$ =0.4; and in a state of 'systemic pessimism,' they assume that $p^* = \overline{p}$ =0.4, and $1-p^* = 1-\overline{p}$ =0.6.

Table 7.7 Equilibrium number of firms

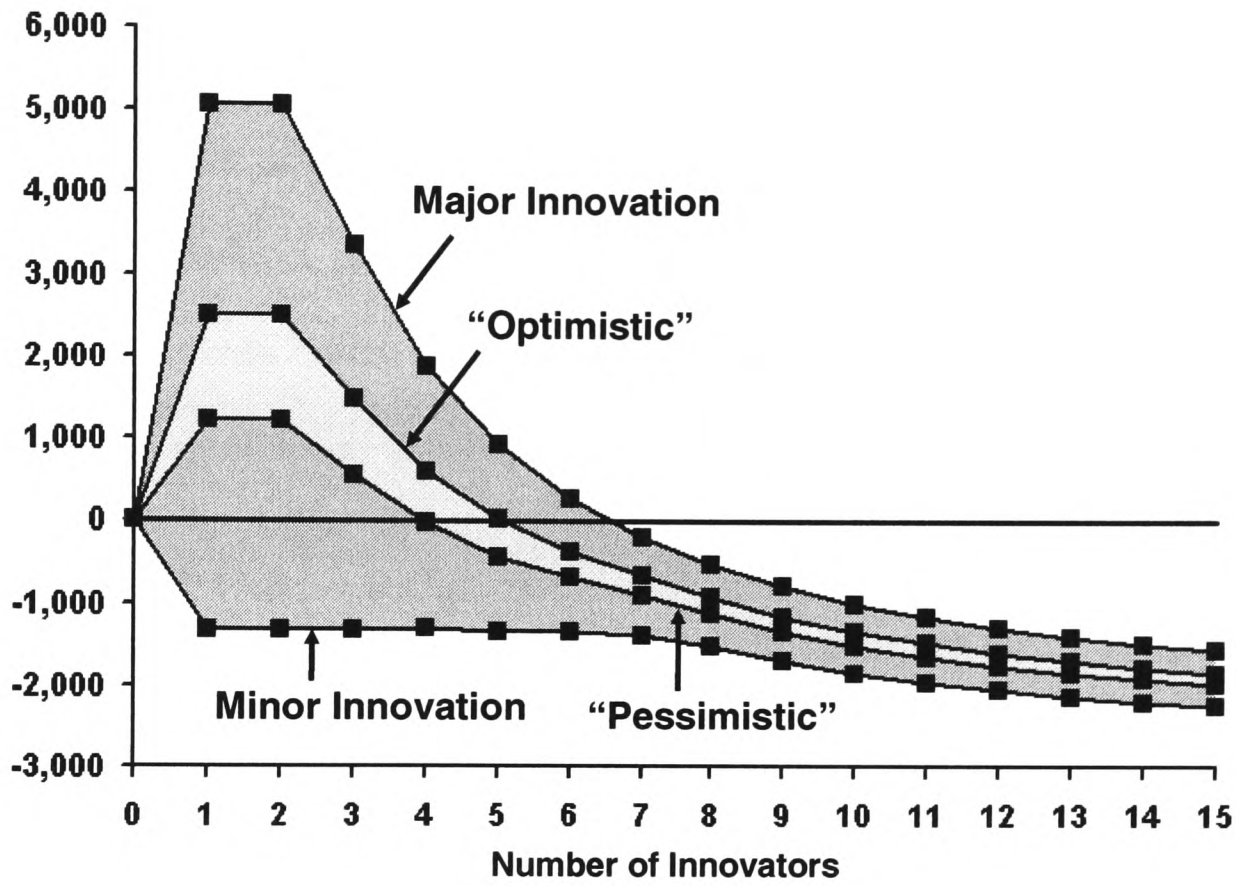
	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Optimistic scenario						
Traditional firms	19	5	0	19	0	0
Innovators		5	5		5	5
Imitators			3			10
Pessimistic scenario						
Traditional firms	19	10	0	19	0	0
Innovators		3	3		3	3
Imitators			5			12

The most significant results corresponding to this example are summarised in Tables 7.7 and 7.8.[9] Table 7.7 shows the extent to which optimism raises the equilibrium number of innovating firms that can expect to get zero expected profits: five firms would pay the sunk cost F_I if they are optimistic about a getting a major innovation (p^*=0.6), against three in a pessimistic state (p^*=0.4). This variation is suggestive of the closeness of the connection between industry-wide spending on innovation and companies' attitude towards fundamental uncertainty in the present approach. And a more optimistic disposition towards uncertainty by the business sector will regulate more than just the aggregate scale of investments in innovative-R&D: it will exert an indirect but significant influence on the entire evolution of the industry. The reason is that entry by a greater number of innovating firms creates a more competitive climate for traditional manufacturers and potential imitators, whatever the

[9] The results from a comparative dynamics analysis exploring the whole range of admissible parameter values – i.e., for values of p* going from 0 to 1 – is reproduced in Appendix 7.2.

outcome of innovation. This intensifies the shakeout of the traditional product, and deters the incursion into the market of product clones. Notice, however, that the *total* number of firms manufacturing the improved design in stage 3 is the same in both cases. This is because, as we showed formally in Chapter 6, there is a one-to-one trade-off between the equilibrium number of innovating and imitating firms in the third stage.

Figure 7.2 Expected and actual profit curves for a typical innovator



In order to clarify the three-way connection between the subjective attitudes to uncertainty, the equilibrium number of innovating firms and these firms' profitability after uncertainty has dissipated, Figure 7.2 plots a typical innovator's (cumulative and discounted) expected and actual profits as a function of the number of innovators in the industry. The topmost boundary in this figure measures the profit each innovator will get when innovation has resulted in a major improvement over the standard product, while the bottommost boundary indicates the profits corresponding to a minor innovation. These are the profits anticipated by a typical innovator after it has factored in the future competition from traditional manufacturers and

imitators in each state of the world. The interior lines correspond to the expected profits when firms are optimistic and pessimistic. The equilibrium number of innovators can be read directly off this figure by moving along the expected profit line up to the point where this drops to zero. Mirroring the results in Table 7.7, this shows that up to three innovators expect to earn nonnegative profits when firms are pessimistic ($p^* = 0.4$); when firms are confident of getting a major innovation ($p^* = 0.6$), the expected profit schedule shifts up, and these beliefs support the entry of five innovators in the industry.

Although the equilibrium level of investment in innovative-R&D is quite sensitive to the prevailing attitude about uncertainty, innovators' *ex post* profitability need not be. Whether or not firm-level profits/losses change on account of there being more or less innovating firms depends on the outcome of the innovation process. Specifically, Figure 7.2 indicates that the equilibrium number of innovating firms has a marked effect on their eventual profit when a major innovation is developed, but no effect whatsoever if the outcome is a minor innovation. The three 'pessimistic' firms investing in innovative-R&D get around 3,500 in profits each when they market a major innovation, whereas the five 'optimistic' firms receive in the order of 1,000. In contrast, the loss made by three or five innovating firms marketing a minor innovation is the same, about 1,300. These findings can be explained by referring back to the formal results of Chapter 6. There it was shown that, conditional on the outcome of the innovation process, a typical innovator's profit would remain unchanged to the extent that additional investments by innovating firms displaced but did not completely drive out traditional and imitating firms. In as much as the increase in the number of innovators brought about a partial displacement of the other producers, the competitive pressure on each innovator's revenue would be insignificant, and each would record the same profit/loss margins; this corresponds to the flat segments in the profit curves observed in Figure 7.2.

But if the number of innovating firms and the outcome of their research efforts was such that traditional and/or imitating firms were 'blockaded' from the market at either the second or third stage, the entry of one more innovator would greatly reduce the revenue and hence profits of each innovator; that is why the profit curve exhibits a 'kink' and is then decreasing in the total number of innovating firms. Average revenue and profits would be smaller because further investments in innovative-R&D have a strong competitive effect: each additional innovator is, from the point of view of the other firms in the industry, an extra competitor with a matching product to contend against in future stages.

On the face of it, the pattern of industry-level profitability that emerges from Table 7.8 is analogous to that experienced by a typical innovator. But it would be a mistake to attribute the second stage slump observed at the aggregate level to the negative competition effect identified as the culprit of the downturn in firm-level profits.[10] The difference in aggregate profitability is driven not by the decline in revenue experienced by individual innovators but by the duplication of the sunk cost of innovation. In this concrete example, duplicated R&D costs amount to 5,000, the value of the investment financed by two (additional) innovating firms. A calculation of a typical innovator's revenue and that corresponding to all innovating firms can ascertain that, even though each innovator earns less revenue, there are more of them, and the combined effect on aggregate revenue is actually positive. With a minor innovation, a typical innovator's revenue in stage 2 is 4,828 in the pessimistic scenario, and 4,917 in the optimistic one, but the revenue across all innovators is 14,750 in the former case, and 24,141 in the latter. In a setting with a major innovation, the revenue for one innovator is 22,222 in the pessimistic scenario and 15,625 in the optimistic one, whereas aggregate revenue is 66,667 under pessimism, 78,125 under optimism. The

[10] When firms have high expectations *ex ante*, industrial profits fall more abruptly in stage 2, whatever the outcome of innovation; there is some small compensation in stage 3, as each of the innovating firms receives as its equilibrium rent the value of the sunk cost required to copy their products.

firm- and industry-level revenue generated by the innovating firms in equilibrium move in opposite directions as beliefs change.

Table 7.8 Aggregate measures of profitability and welfare

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Optimistic scenario						
Industrial profits	63	– 8,887	2,865	63	2,625	2,759
Consumer welfare	9,814	10,514	11,622	9,814	23,438	37,037
Social welfare	9,877	1,627	14,487	9,877	26,063	39,796
Pessimistic scenario						
Industrial profits	63	– 5,158	1,865	63	8,867	1,759
Consumer welfare	9,814	10,142	11,622	9,814	16,667	37,037
Social welfare	9,877	4,984	13,487	9,877	25,533	38,796

So far, it has been suggested that a business sector that is confident about the prospects of innovating will finance a greater number of costly investment projects, which in some circumstances can lead to more aggressive rivalry in the final product market, hurting the finances of individual innovators – this occurs if the negative competition effect is binding, which is more likely when a major innovation has been developed. In addition, the profitability of the business sector as a whole is curtailed by the duplication of R&D costs. The results would appear to support the widespread view that optimism can give rise to excess investment in innovative activities and to overproduction. But there are at least two reasons why it would be wrong to label the proliferation of R&D as "excess investment." First, the number of projects financed in the industry is determined precisely by the anticipation of zero expected profits. As a result, the notion of excess or under-investment is not especially meaningful until after the uncertainty has dissipated.

Second, the losses sustained by firms in a more competitive industry are offset by a gain to consumers, as they are offered greater quantities of the new good at lower prices. The table above shows that, whatever the outcome of innovation, consumer welfare is greater in each stage when firms are more optimistic, and the gains are substantial when the firms are

selling a greatly enhanced commodity. The net outcome of this trade-off between industrial profits and consumer welfare is not at all clear. Social welfare increases in stage three when firms are more optimistic, as the sum of innovators' quasi-rents increases, but in the second stage welfare may increase or decrease, and what transpires depends on the success of own-product R&D. The opposing impacts on social welfare make a comprehensive assessment of a particular stance towards uncertainty much more difficult.

One important aspect of such an evaluation would be whether *any* innovation is financed when firms are less optimistic. Given the other parameter values in the model, the range of uncertainty explored by these two simulations always results in some firms disbursing the fixed cost required for innovation. Notice in Figure 7.2 that the expected profit schedule for any probability p^* in the range $[.4,.6]$ will be located within the limits of the lightly shaded area. As a result, firms' beliefs about uncertainty cannot lead to a stagnant, non-innovating industry. But when the range of uncertainty is broader, firms' disposition towards it – their confidence, 'naïve optimism,' or 'animal spirits,' to use Keynes's expression – could become the crucial factor determining whether or not any product development is financed. In Table A7.1 (in the Appendix), the equilibrium number of innovating firms drops to zero if the subjective probability of a major innovation falls below p^*=0.3. Naturally, the sway of uncertainty over investment decisions will be all the stronger if the cost structures and market conditions depress the profit schedules for a minor and/or a major innovation – e.g., when the fixed cost of production is lower, which would induce relatively more competition from traditional manufacturers, when firms apply a stronger discount to future payoffs, when imitation is less costly, etcetera.

On the other hand, we should emphasise that the influence of the subjective probability estimates on the sustainable level of investment in innovative-R&D is capped from above in the solution to the model. The results in Table A7.1 indicate that (given the

value of the other parameters,) the equilibrium number of innovators will never be above six, no matter how confident firms are about their research efforts. This can be observed from Figure 7.2 as well: in the extreme case where firms are certain about getting a major product innovation ($p^* = 1$), the expected profit schedule would coincide with the uppermost boundary in the figure, which drops below the zero profit line if there are more than six firms investing in innovation. The upper bound to aggregate R&D of the innovative kind is a direct consequence of the forward-looking nature of firms' estimates of the profit streams that are likely to be generated by this research activity. The ceiling to innovative-R&D can be determined from the cost structure in the industry, the market demand function, the type of product market competition, and the anticipated outcomes of the innovation process.

Note, by contrast, that there is no such ceiling to aggregate spending on imitative-R&D. The sustainable entry of imitating firms in the model depends on research and production costs, demand, and competitive conditions, but also on the aggregate (equilibrium) level of innovative-R&D investment and, crucially, on the realised outcome of these investments. For simplicity, this chapter has compared the industry evolutions in which the realised outcomes of the innovation process turn out to be the same as the anticipated outcomes. But these are just two of the possible outcomes that can be supported in the game-theoretic model we have been studying. As the previous chapter and the introduction to this one explained, the realised outcome does not have to coincide with the outcomes envisioned *ex ante*. It does not even have to lie within the set enclosed by these outcomes: the innovating firms may end up with a product design that is only infinitesimally better than that of the conventional product, or with one that far surpasses their original ambitions. As a consequence, the number of firms that eventually seek to imitate the new product design can be far lower or greater than what the industry had originally foreseen.

7.4 Conclusions

In this chapter, numerical methods have been employed to simultaneously determine the industry structure and the aggregate scale of investments in innovative- and imitative-R&D within a dynamic, uncertain setting. By examining the relationship between innovators and imitators, the analysis has aimed to capture some key aspects of Schumpeter's 'vision' of the dynamics of technologically-driven development. Schumpeter's view that the prospect of monopoly mark-ups are required to induce fixed R&D investment is seen to be valid – so long as we do not view it as the long-run equilibrium of the market when innovation is successful and most disruptive. In the latter case, the transient monopoly power before the "entry of the clones" is sufficient to induce innovators. Considered in this light, 'creative destruction' is not an incremental process, such as envisioned in the Aghion-Howitt equilibrium of a simple representative firm. What Aghion and Howitt (1992) term the 'creative destruction' effect is narrowly circumscribed to the reduced incentives towards current research corresponding to the chance that a new innovation will come along, thus eroding the profits of the current patent-holders. The 'process of industrial mutation' is therefore reduced to the replacement of one monopolist for another that has developed an improved good.[11]

Our theoretical approach leads to a different appreciation of the dynamics of technological change. In it, the shakeout of incumbent firms is as much a consequence of successful technological breakthroughs by outsiders as of the cheaper variants of the new product supplied by imitating firms. In terms of the metaphor coined by Schumpeter, one could say that the "wave of innovation" may allow some of the traditional firms to keep themselves above water, but when it induces a second "wave of imitation" the standard

[11] See the third section of Chapter 2 for a discussion of the merits and shortcomings of Aghion and Howitt's formalisation of Schumpeter's theory of economic change.

product will be washed away from the market. But, seen from the viewpoint of those investing in innovative-R&D, a major innovation, although it is very successful in dislodging the established products, will also spur a "swarm of imitators" into entering the industry. Thus, the very success of an innovator's research activities creates openings for competition from rival firms, ultimately bringing about a sharp reduction in its price mark-ups and profits.

Although the general tenor of the chapter is evocative of the evolutionary theory of "Schumpeterian competition" *a la* Nelson and Winter (1982, 2002), this is not an evolutionary model. There are no bounds on the firms' "information-processing capacities": each firm selects its output level to maximise profits given the prevailing industry composition and production costs, and the investment decisions in the industry fully anticipate the type and extent of competition that firms could encounter in the future. But, while this is "basically a maximization-and-equilibrium kind of" model, as Paul Krugman would put it,[12] innovators' profit maximisation is subject to the constraints imposed by Knightian uncertainty regarding the outcome of research. The model sought to formalise two of the ideas about risk and uncertainty that Knight reiterates in his 1921 treatise: first, that there is a gradation between the two forms of randomness; and, second, that more accurate knowledge or a better judgement by the agent(s) in charge of the investment decisions can moderate but not eliminate fundamental uncertainty.[13] Concretely, firms decided their

[12] Paul Krugman, "What Economists Can Learn From Evolutionary Theorists," A talk given to the European Association for Evolutionary Political Economy, November 1996, at www.mit.edu/~krugman/evolute.html.

[13] Uncertainty, Knight (1957, p. 346) wrote: "is as ineradicable from business decisions as from those in any field. The amount of uncertainty may, however, be reduced in several ways, as we have seen. In the first place, we can increase our knowledge of the future through scientific research and the accumulation and study of the necessary data." Now, any simple definition of uncertainty is bound to lose some of the nuances of Frank Knight's thought on this phenomenon. Some among the unmodelled aspects of his discourse on uncertainty are summarised in the following lines, where he mentions three more methods that can reduce uncertainty. One is "by the clubbing of uncertainties through large-scale organization of various forms." Another, by increasing "control over the future" through restrictions on the freedom of some agents in the economy. Lastly, "by slowing up the march of progress."

funding of innovative-R&D in a setting of "imprecise risk," which led to their point-estimates of the probabilities of developing a given innovation being partly based on subjective factors.

In the solution to the model, the actual sequence of R&D and entry in a sustainable equilibrium reflects any divergence between the *ex ante* beliefs concerning the uncertainty surrounding technological advances and the realised outcomes of the innovation process. This implies that, in almost all cases, innovators' realised profits would not coincide with their expected value of zero. As a rule, innovating firms end up with a loss or a profit.[14] Section 7.3 was able to further precise the relation between uncertainty, beliefs and profits. A favourable disposition towards uncertainty was consistent with greater investment in innovative-R&D and, as a result, traditional and imitating firms were "crowded out" from the final product market. The strength of the connection between firms' *ex ante* beliefs and innovators' *ex post* profitability was seen to depend on whether the innovating firms partially displaced or completely blockaded the traditional and/or imitating firms. Even if the conditions are such that firm-level profits are unaffected by the equilibrium number of innovating firms, additional R&D duplication would still harm the industry-level measures of profitability. Not all was bad, however, as the firm's expected profit maximisation did not take into account of the gains to consumers from having more rivalry in the market. Although consumers benefited from the increased competitiveness of the industry, it was unclear whether social welfare would be higher or lower.

[14] The Knightian 'flavour' of this result is inescapable. Frank Knight's intuition of the relationship between uncertainty and profit was that: "Thus uncertainty explains profits and loss; but profit, when it occurs, is not properly a "reward for risk taking," though the expectation of gain is the incentive for assuming the entrepreneurial role" (Preface to the 1957 reprint, p. lxii). Knight (1957, pp. 310-1) also makes the following comment: "The only "risk" which leads to profit is a unique uncertainty resulting from an exercise of ultimate responsibility which in its very nature cannot be insured nor capitalized nor salaried. Profit arises out of the inherent, absolute unpredictability of things, out of sheer brute fact that the results of human activity cannot be anticipated and then only in so far as even a probability calculation in regard to them is impossible and meaningless."

As a first approximation to this problem, the analysis in these chapters assumed that the firms had identical and exogenous beliefs about the probability of developing a given product improvement. The last section of Chapter 6 suggested that process of belief formation could be studied by introducing an additional stage into the model. But if one is content to rely on numerical methods, there is an alternative that is conceptually more attractive. This is to repeat the sequential game, with its basic innovation-imitation structure, and allow beliefs to change from one cycle of structural change to the next, according to the realised outcome of the innovation process. To put this idea into practice, one would need to specify the stochastic process that controls the outcomes from the investments in innovative-R&D. Then beliefs could evolve as follows. When innovating firms obtained positive profits as a result of their successful research efforts, they would become more optimistic about the future likelihood of finding a major innovation. If, on the contrary, they ended up with a net loss, they would become more pessimistic about their research capabilities in the next cycle of "creative destruction."[15]

The equilibrium concept we have introduced determines R&D spending and market structure endogenously. In this respect, the analysis "stands on the shoulders" of the theoretical contributions of Partha Dasgupta and Joseph Stiglitz, who, in 1980, showed how innovation and market structure could be simultaneously determined from underlying conditions (especially firm or product market characteristics). Up to that date, the relationship between innovation and market structure had been thought of in terms of lines of causality; that is, whether firms with given market shares were more likely to innovate or, conversely, if innovation tended to make industries more concentrated. Although there was no causality in

[15] This dynamic structure is likely to create a degree of persistence in the investments directed towards innovation over time, generating more elaborate patterns of industrial innovation and growth. Naturally, one would want to circumscribe the degree of optimism or pessimism to the objective bounds established by the firms on the basis of the knowledge available to them, which would tend to dampen the dynamics of this extended process of technological change.

Dasgupta and Stiglitz's (1980a) analysis, they concluded that a positive statistical relationship would still be observed between industrial concentration and R&D intensity in a cross-section of industries. The results in this chapter suggest that innovative-R&D investments would be positively related to industrial concentration in the short run, but not always in a longer time frame that encompasses the entry of imitating firms into the industry. Indeed, in industries where major product innovations have been developed, industrial measures of concentration may follow an inverted U-shaped path.[16]

Another particularity of our equilibrium concept is that, by focusing on completely symmetric situations, it does not predict whether innovating and imitating firms will mostly be incumbent firms already producing the conventional product, or new entrants to the industry. Without adding specific, heterogeneous characteristics of firms that would account for the identities of the innovators, entering imitators, and the firms that were forced out of the industry, the historical details will remain unexplained. One should note that some recent analyses of technological evolution have taken the opposite tack. The so-called "history-friendly" models, the newest 'generation' of evolutionary models, are structured so as to fit the historical data of specific industry evolutions. In the words of the original contributors to this tradition, "[h]istory-friendly models are formal models that aim to capture, in stylized form, qualitative and 'appreciative' theories about the mechanisms and factors affecting industry evolution, technological advance and institutional change put forth by empirical scholars of industrial economics, technological change, business organization and strategy, and other scientists" (Malerba et al 1999, pp. 3-4).[17] But one could argue that it is more

[16] This is because, as we already noted, a radically improved product initially brings about a rapid shakeout of traditional producers, fostering a less competitive market characterised by high price mark-ups and profit margins. But the exceptional profitability of major improvements tends to induce other firms to clone the new design, the equilibrium price of the improved good will be driven down and profit margins in the industry should go back to normal levels.

[17] For the origins of history-friendly models and some applications to the computer industry, see Malerba et al (1999, 2001a,b). One of the attractive aspects of tailor-making the models is that

important from the *social viewpoint* to know how many innovators and competitors will emerge and survive, than it is to be able to predict their identities, however interesting such historical details may be to the private investors and employees involved.

The numerical results presented in this chapter suggest that the combined effect of successful innovation and intense imitation is to greatly increase social welfare. In those circumstances, the industry is profitable, and consumers benefit on two counts: they can purchase a good of much higher quality than was previously available; and the price of this good, although it is quite high during the stage in which only innovators can market it, falls sharply as soon as "product clones" are commercialised by imitators. Conversely, unsuccessful innovation (relative to the industry's original expectations) and slight imitation can result in welfare decreasing below the levels predicted for a static, non-innovating industry. The latter result is partly accounted by the irreversibility of research investments, as investments in innovative-R&D are permanently lost to the industry. Consumers are also burdened by these errors of judgement, as they face a progressively more concentrated market structure and higher prices in exchange for a small increase in product quality. The purpose of the next (and last) chapter is to investigate whether public policies aiming to encourage greater competitiveness in the final product market, or a different balance of innovative/imitative activities, can increase consumer and/or social welfare in this framework, and the implications of so doing for the evolution of the industry equilibrium.

---

counterfactuals can be introduced, and the predicted evolution computed. For instance, one of the dominant firms in, say, the PC industry, can be split in two, and the effects of this on future market structure can be predicted (under the assumption that firms will not change their behaviour in response to the change in industry structure).

Appendices

Appendix 7.1 Parameter values used in the simulations

Parameter	Benchmark	Optimistic Firms	Pessimistic firms
S	100,000	100,000	100,000
δ	0.8	0.8	0.8
σ	1	1	1
c	0.5	0.5	0.5
Z_T	1	1	1
Z_L	1.25	1.25	1.25
Z_H	2	2	2
F_P	100	100	100
F_I	2500	2500	2500
F_C	500	500	500
p^*	0.5	0.6	0.4

Appendix 7.2 Comparative Equilibrium Dynamics Results

With respect to the estimated probability of (a major) innovation (p^*)

Table A7.1 Number of firms in 'sustainable equilibrium'

	Minor Innovation					Major Innovation				
p^*	Traditional firms			Innovative firms	Imitating firms	Traditional firms			Innovative firms	Imitating firms
	Stage 1	Stage 2	Stage 3			Stage 1	Stage 2	Stage 3		
0	19	19	19	0	0	19	19	19	0	0
0.1	19	19	19	0	0	19	19	19	0	0
0.2	19	19	19	0	0	19	19	19	0	0
0.3	19	10	0	3	5	19	0	0	3	12
0.4	19	10	0	3	5	19	0	0	3	12
0.5	19	7	0	4	4	19	0	0	4	11
0.6	19	5	0	5	3	19	0	0	5	10
0.7	19	5	0	5	3	19	0	0	5	10
0.8	19	5	0	5	3	19	0	0	5	10
0.9	19	2	0	6	2	19	0	0	6	9
1	19	2	0	6	2	19	0	0	6	9

Table A7.2 Indicators of market structure

p^*	Minor Innovation						Major Innovation					
	Total number of firms			Herfindahl-Hirschman Index			Total number of firms			Herfindahl-Hirschman Index		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0	19	19	19	526	526	526	19	19	19	526	526	526
0.1	19	19	19	526	526	526	19	19	19	526	526	526
0.2	19	19	19	526	526	526	19	19	19	526	526	526
0.3	19	13	8	526	1,108	1,250	19	3	15	526	3,333	667
0.4	19	13	8	526	1,108	1,250	19	3	15	526	3,333	667
0.5	19	11	8	526	1,260	1,250	19	4	15	526	2,500	667
0.6	19	10	8	526	1,316	1,250	19	5	15	526	2,000	667
0.7	19	10	8	526	1,316	1,250	19	5	15	526	2,000	667
0.8	19	10	8	526	1,316	1,250	19	5	15	526	2,000	667
0.9	19	8	8	526	1,429	1,250	19	6	15	526	1,667	667
1	19	8	8	526	1,429	1,250	19	6	15	526	1,667	667

Table A7.3 Price mark-ups

p^*	Minor Innovation					Major Innovation				
	Standard product			Improved product		Standard product			Improved product	
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3
0	9 %	9 %	9 %	---	---	9 %	9 %	9 %	---	---
0.1	9 %	9 %	9 %	---	---	9 %	9 %	9 %	---	---
0.2	9 %	9 %	9 %	---	---	9 %	9 %	9 %	---	---
0.3	9 %	9 %	---	21 %	18 %	9 %	---	---	33 %	11 %
0.4	9 %	9 %	---	21 %	18 %	9 %	---	---	33 %	11 %
0.5	9 %	10 %	---	21 %	18 %	9 %	---	---	29 %	11 %
0.6	9 %	9 %	---	21 %	18 %	9 %	---	---	25 %	11 %
0.7	9 %	9 %	---	21 %	18 %	9 %	---	---	25 %	11 %
0.8	9 %	9 %	---	21 %	18 %	9 %	---	---	25 %	11 %
0.9	9 %	9 %	---	21 %	18 %	9 %	---	---	22 %	11 %
1	9 %	9 %	---	21 %	18 %	9 %	---	---	22 %	11 %

Table A7.4 Output levels

p^*	Minor Innovation					Major Innovation				
	Traditional firm			Improved product		Standard product			Improved product	
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3
0	2,273	2,273	2,273	---	---	2,273	2,273	2,273	---	---
0.1	2,273	2,273	2,273	---	---	2,273	2,273	2,273	---	---
0.2	2,273	2,273	2,273	---	---	2,273	2,273	2,273	---	---
0.3	2,273	2,344	---	8,138	7,102	2,273	---	---	33,333	11,111
0.4	2,273	2,344	---	8,138	7,102	2,273	---	---	33,333	11,111
0.5	2,273	2,381	---	8,185	7,102	2,273	---	---	28,571	11,111
0.6	2,273	2,244	---	8,013	7,102	2,273	---	---	25,000	11,111
0.7	2,273	2,244	---	8,013	7,102	2,273	---	---	25,000	11,111
0.8	2,273	2,244	---	8,013	7,102	2,273	---	---	25,000	11,111
0.9	2,273	2,273	---	8,049	7,102	2,273	---	---	22,222	11,111
1	2,273	2,273	---	8,049	7,102	2,273	---	---	22,222	11,111

Table A7.5 Firm-level profit/loss margins

p^*	Minor Innovation						Major Innovation					
	Traditional firms			Innovative firms		Imitating firms	Traditional firms			Innovative firms		Imitating firms
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3		Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	
0	0 %	0 %	0 %	---	---	---	0 %	0 %	0 %	---	---	---
0.1	0 %	0 %	0 %	---	---	---	0 %	0 %	0 %	---	---	---
0.2	0 %	0 %	0 %	---	---	---	0 %	0 %	0 %	---	---	---
0.3	0 %	1 %	---	– 36 %	13 %	1 %	0 %	---	---	13 %	8 %	0 %
0.4	0 %	1 %	---	– 36 %	13 %	1 %	0 %	---	---	13 %	8 %	0 %
0.5	0 %	1 %	---	– 35 %	13 %	1 %	0 %	---	---	8 %	8 %	0 %
0.6	0 %	0 %	---	– 37 %	13 %	1 %	0 %	---	---	3 %	8 %	0 %
0.7	0 %	0 %	---	– 37 %	13 %	1 %	0 %	---	---	3 %	8 %	0 %
0.8	0 %	0 %	---	– 37 %	13 %	1 %	0 %	---	---	3 %	8 %	0 %
0.9	0 %	0 %	---	– 36 %	13 %	1 %	0 %	---	---	– 1 %	8 %	0 %
1	0 %	0 %	---	– 36 %	13 %	1 %	0 %	---	---	– 1 %	8 %	0 %

Table A7.6 Industry-wide sales and profitability

p^*	Minor Innovation						Major Innovation					
	Sales Volume			Industrial Profits			Sales Volume			Industrial Profits		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0	23,554	23,554	23,554	63	63	63	23,554	23,554	23,554	63	63	63
0.1	23,554	23,554	23,554	63	63	63	23,554	23,554	23,554	63	63	63
0.2	23,554	23,554	23,554	63	63	63	23,554	23,554	23,554	63	63	63
0.3	23,554	27,568	33,574	63	– 5,158	1,865	23,554	66,667	92,593	63	8,867	1,759
0.4	23,554	27,568	33,574	63	– 5,158	1,865	23,554	66,667	92,593	63	8,867	1,759
0.5	23,554	28,926	33,574	63	– 6,877	2,365	23,554	73,469	92,593	63	5,927	2,259
0.6	23,554	30,254	33,574	63	– 8,887	2,865	23,554	78,125	92,593	63	2,625	2,759
0.7	23,554	30,254	33,574	63	– 8,887	2,865	23,554	78,125	92,593	63	2,625	2,759
0.8	23,554	30,254	33,574	63	– 8,887	2,865	23,554	78,125	92,593	63	2,625	2,759
0.9	23,554	31,603	33,574	63	– 10,617	3,365	23,554	81,481	92,593	63	– 785	3,259
1	23,554	31,603	33,574	63	– 10,617	3,365	23,554	81,481	92,593	63	– 785	3,259

Table A7.7 Consumer and social welfare

p^*	Minor Innovation						Major Innovation					
	Consumer Welfare			Social Welfare			Consumer Welfare			Social Welfare		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0	9,814	9,814	9,814	9,877	9,877	9,877	9,814	9,814	9,814	9,877	9,877	9,877
0.1	9,814	9,814	9,814	9,877	9,877	9,877	9,814	9,814	9,814	9,877	9,877	9,877
0.2	9,814	9,814	9,814	9,877	9,877	9,877	9,814	9,814	9,814	9,877	9,877	9,877
0.3	9,814	10,142	11,622	9,877	4,984	13,487	9,814	16,667	37,037	9,877	25,533	38,796
0.4	9,814	10,142	11,622	9,877	4,984	13,487	9,814	16,667	37,037	9,877	25,533	38,796
0.5	9,814	10,240	11,622	9,877	3,363	13,987	9,814	20,408	37,037	9,877	26,335	39,296
0.6	9,814	10,514	11,622	9,877	1,627	14,487	9,814	23,438	37,037	9,877	26,063	39,796
0.7	9,814	10,514	11,622	9,877	1,627	14,487	9,814	23,438	37,037	9,877	26,063	39,796
0.8	9,814	10,514	11,622	9,877	1,627	14,487	9,814	23,438	37,037	9,877	26,063	39,796
0.9	9,814	10,619	11,622	9,877	1	14,987	9,814	25,926	37,037	9,877	25,141	40,296
1	9,814	10,619	11,622	9,877	1	14,987	9,814	25,926	37,037	9,877	25,141	40,296

Chapter 8

The potential for welfare-enhancing competition and innovation policies

8.1 Introduction

The numerical analysis in chapter 7 indicated that major success in the innovation process appears to bring substantial benefits for consumers in two ways: first, the initial rise in consumer welfare stemming from purchases of the new, higher quality goods of the innovating firms, and second, a subsequent gain in consumer surplus when the entry of a "swarm of imitators" drives down the prices of those goods. Alternatively, a minor improvement in quality would bring only marginal welfare gains to consumers – even when product clones become available. This is because the industry becomes progressively more concentrated and prices are adjusted upwards to extract their "willingness to pay" for the incrementally better product quality.

The theoretical basis for these patterns lies in the simultaneous determination of the investments in innovative- and imitative-R&D, the quality of the goods offered at each stage, and the evolving market structure. Although the scale of the investments, the goods' quality and the market structure are interlocked in the solution of the model, it is convenient to think of the research strategies as having two distinct effects on consumer welfare. A positive one arising from the enhanced quality of the goods consumers can buy, first supplied by innovating firms alone and later on by imitating firms also. And another, which can be positive or negative, as the introduction of an improved product line changes the industry structure; the sign of this effect depends on the precise scale of the investments made by

innovating and imitating firms, as this controls the pace at which the traditional product line is driven from the market.

This conceptual division may be fruitful in pointing to two classes of policy instruments that could be employed to improve the dynamic system's performance in welfare terms. One class includes different types of *innovation policies* that intend to shift the balance of R&D investments (in equilibrium) by altering the cost of new product development vis-à-vis imitation. The tax-transfer system is the most important example of an economic 'lever' that can be used to push the relative cost of innovative- and imitative-R&D one way or another; statutory and judicial provisions affecting patent protection, and the exceptions available for "reverse engineering" are the chief legal instruments the government has to ease or put hurdles to imitation. The second class of policy instruments aims to promote pro-competitive behaviour by the firms in the industry and to rectify any anti-competitive tendencies that seem endemic to the process of structural change. These can be generically denoted as *competition policies*.[1] Both sets of policy instruments are to be distinguished from 'command-and-control mechanisms,' which will not be treated in this chapter.[2] The central question which this chapter addresses in the context of the modelling approach developed in Chapters 6 and 7, is whether, and under what conditions these two classes of public policy

[1] The theoretical and empirical literature on the set of policies we have listed is far too extensive for this chapter to summarise its main results. There are, however, a number of key references, as well as special journal issues on these subjects; these contain a more extensive list of citations to this literature. Concerning innovation policies, see Stoneman and Dierden (1994), Metcalfe (1995), Metcalfe and Georghiou (1998), Carlsson and Jacobsson (1997), Lipsey (1998), Malerba (1998), OECD (1998) and Hall (2002). Also the articles in the May 1983 issue of the *American Economic Review*, Volume 73, which was devoted to "R&D and Productivity Increases"; those in the May 1986, Volume 76, issue of the same journal, addressing "R&D, Innovation, and the Public Policy"; and the Spring 1990 issue of the *RAND Journal of Economics*, Volume 21, that collects the papers from a "Symposia on Patents and Technology Licensing." For discussions of evolving EU innovation policies see Soete and Arundel (1993) and Cowan and van de Paal (1999). With respect to competition policies and innovation, one should mention Ordover and Baumol (1988), Vickers (1995), Stiglitz (1997), Martin (2000), and the collection of articles in the Summer 1990 issue of the *Journal of Economic Perspectives*, Volume 4, on "Collaboration, Innovation and Antitrust."

[2] Compulsory licensing of innovations, price controls, and divestiture (break-up of a dominant firm) are examples of command-and-control mechanisms.

tools could be employed to increase expected social welfare. The analysis will also establish the consequences of the use of these policy instruments for product development, the viability of the conventional product line, and the resulting evolution in pricing, firm growth, and industrial profitability.

The order of the chapter, its methods and central results are as follows. In the second section, we examine the relationship between product market competition, R&D spending and the pattern of industrial dynamics. It is clear that the profits that firms expect to earn by investing in innovative- or imitative-R&D, or by staying put and manufacturing the conventional product line, depend negatively on the intensity with which they subsequently will compete with each other for sales. It follows that the toughness of price competition is a critical determinant of the equilibrium number of firms of each kind. What is not immediately obvious is whether a particular form of competition is relatively more favourable to innovators, given their reliance on the revenue collected by competing with a better product before it becomes widely available to competitors? Or whether it is the producers of the conventional product lines that stand to gain, allowing them to stay afloat for longer? And, whatever the relative bias towards a specific research activity or product line from a certain mode of rivalry, how is market structure and social welfare affected? These questions are addressed through a two-pronged approach, which (i) compares the numerical solution with quantity competition in the final product market to that in which there is competition in prices, and (ii) calculates the comparative equilibrium dynamics with respect to the degree of product substitutability – ranging from markets in which firms have niches that are completely protected from other producers, to those where there is perfect horizontal substitution between products.

The simulation results confirm that more intense competition acts as a constraint on industry-wide spending on innovative- and imitative-R&D and on the entry of traditional firms; this is true of the Bertrand solution and in the analysis of markets with less horizontal differentiation. As a result of the weak incentives to fund both classes of R&D or produce the conventional product, the industry becomes more concentrated when rivalry is more acute. The trade-off between (endogenous) concentration and (exogenous) competition severely erodes the welfare gains that one might have anticipated from a setting where firms compete more vigorously. These results have interesting (quite Schumpeterian!) implications for the conduct of competition policy: pro-competitive policies may be ineffective or even counterproductive when firms need to anticipate an ability to fund substantial sunk costs in order to generate technological advances. In other words, there is a tension between static allocative efficiency and the dynamic processes of knowledge creation and diffusion that have the potential to raise consumer welfare over time. By the same token, it is no longer clear that structural measures of market power reveal much information about the actual competitiveness of an industry. There is separate problem with the measurement of market power for the purposes of implementing competition policy: because such measures are calculated at one point in time, they do not reveal whether a firm or firms that currently profit from an existing technological lead are likely to lose their dominance quickly to firms that could enter the market by investing in imitative-R&D.

Section 8.3 turns to analyse the benefits of innovation policies intending to tip the balance between innovative and imitative efforts. The section begins by calculating the social welfare maximising number of innovative firms and comparing it with the symmetric sustainable equilibrium. The difference between the two is found to be negligible. We then present the results from a comparative dynamics analysis with respect to the sunk cost of imitation. The simulations suggest that overall consumer welfare would be improved by a

policy facilitating imitation. But this result masks a redistribution of industrial profitability and consumer welfare across the last two stages of the industry evolution. Welfare would be higher in the third stage because additional entry by imitating firms leads to a much more competitive industry. That prospect, however, must have an adverse effect on the investments in innovative-R&D: firms anticipate that there will be pervasive imitation, reducing the rent stream from any given innovation, and hence there are less innovators in equilibrium; consequently, in the realisation of the second stage, the industry would be more concentrated than otherwise, and consumer welfare would be lower.

The last question we seek to answer is whether such policies seem worthwhile if the government is using transfers financed by general taxes to influence firms' decisions. The answer is negative: social welfare increases would not justify giving a targeted subsidy to imitating firms, and the net social loss (after deducting the transfers) is greater the higher the level of the subsidy. The section concludes that in this game theoretic setting innovation policies have only a limited scope in raising welfare, because the competitive mechanisms are quite effective (by themselves) in curbing the market power of the firms in the industry.

The final section of this chapter steps back from the detailed analysis of policy instruments in order to consider the main policy-related limitations of the theoretical investigations undertaken in this thesis, and the ways in which those shortcomings might be remedied. The resulting agenda for future research is structured along four main axes that collectively aim to bring the models into closer conformity with the various 'milieus' in which firms operate in different countries. Specifically the discussion calls for elaboration of distinctive features of the markets for R&D workers and for innovation finance, extending the study of technology policies, and examining the relation between 'institutional infrastructures' as a whole and structural change in the economy.

8.2 Competition, structural change and the public interest

8.2.1 A comparative analysis of Cournot and Bertrand competition

As we noted in Chapter 6, the specification of demand and the type of competition are two assumptions that can be changed without affecting the dynamic structure (and hence the solution process) in the sequential game theoretic model. It is simply a question of substituting the Bertrand-Nash equilibrium prices, quantities and profits in place of those corresponding to the Cournot-Nash equilibrium, and going through the solution procedure once more. Because the Bertrand equilibrium with a linear demand structure (obtained from the augmented quadratic utility functions) has been derived by Symeonidis (2003), this substitution is readily accomplished. Tables 8.1 to 8.6 reproduce the numerical results already discussed in the second section of Chapter 7, and present those from simulations with price competition in the final product market. To make these results comparable, the parameter values are the same in both cases.

Notice first that all of the firm- and industry-level tendencies that emerged from our analysis under the assumption of quantity competition also are observed in the solution under the assumption of price competition. Spending on innovative-R&D is positive, as are the expenditures on imitative-R&D, which are greater with a major innovation than with a minor one. The industry becomes more concentrated following the entry of innovators. Product cloning worsens this situation when a minor innovation has been developed, but the industry becomes more competitive when a "swarm of imitators" is induced by a major innovation. Similarly, the paths in pricing, firm growth, profit margins, aggregate profitability and welfare parallel those in the numerical solution with firms competing in output. Given the likeness between the firms' equilibrium profit functions in the two cases, the similarity of the numerical results is not surprising. It is, nonetheless, a sign of their robustness to changes in the underlying assumptions, in this case to the type of competition in the final product

market. The following sections further increase our confidence in the claims made (solely) on the basis of a numerical analysis, by showing that the central features of the solution appear over quite a wide range of parameter values.

Naturally, if one examines the solutions closely, differences between the two settings become apparent. In the Bertrand equilibrium solutions there are fewer traditional, innovative and imitative firms in each stage of the game, for both the major and minor innovation outcomes of the R&D process. In the pre-innovation stage, the market supports 14 traditional firms, against the 19 firms that could produce the conventional product in the setting with quantity competition. Then 3 rather than 4 firms invest in product innovation, and the feasible number of traditional firms in stage 2 is lower too. Lastly, under price competition, fewer firms are willing to pay the sunk cost of cloning the new product line; in this case, 3 firms clone the minor product innovation vs. the 4 with quantity competition, and 8 (instead of 11) firms develop and manufacture a product clone of the major innovation. These are natural enough findings when one considers that with this tougher form of product market competition there is less revenue to be gained, and hence the number of firms that can recover the fixed costs of production or research should be lower too.

Table 8.1 Number of firms in the industry

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Cournot competition						
Traditional firms	19	7	0	19	0	0
Innovators		4	4		4	4
Imitators			4			11
Total	19	11	8	19	4	15
Bertrand competition						
Traditional firms	14	5	0	14	0	0
Innovators		3	3		3	3
Imitators			3			8
Total	14	8	6	14	3	11

It would seem that a more acute form of competition is not comparatively more 'damaging' to firms adhering to one or another of the 'innovation strategies,' as there is an almost uniform reduction of the equilibrium entry rates in each stage of the game when prices (rather than quantities) are being set; that is, the ratio of innovators to the initial number of firms, and of imitators to innovators, remains approximately the same in both the Bertrand and Cournot versions of the solution.[3] The upshot is that when competition in the final product market is more acute, the industry may give the impression of being less competitive – at least, if attention is centred on structural measures of market power, an approach favoured by competition authorities in the United States. Whether one takes the total number of firms coexisting in the industry as an inverse measure of concentration, or looks directly at the values taken by the Herfindahl-Hirschman index, the industry is considerably more concentrated in the equilibrium with Bertrand competition; the Herfindahl index takes on values that are approximately 35% higher with price competition.

Table 8.2 Herfindahl-Hirschman index

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Cournot competition	526	1,260	1,250	526	2,500	667
Bertrand competition	714	1748	1667	714	3333	909

That, however, is not the full story: the price mark-ups set on the standard and the improved products, an alternative measure of the firms' effective market power, show an across-the-board reduction. The average mark-up is around 20% lower in each stage when firms are competing in prices rather than quantities, whatever the success of the innovating firms' R&D. The theoretical basis for the observed discrepancy in these measures of market power

[3] To validate these assertions rigorously, one would need to derive the properties of the sustainable entry and profit functions in a "symmetric sustainable equilibrium," and compare these to the ones corresponding to quantity competition in Chapter 6. This, however, is left for future work.

is the following. For a given industrial configuration, the revenue of any firm manufacturing a good of a specific quality is less under Bertrand competition than under Cournot. Therefore, the number of firms at the sustainable equilibrium will necessarily be less in the former case: the industry needs to be more concentrated if each firm is to recover (or expect to recover) its fixed costs. At the same time, the equilibrium price commanded by a specific product is lower with Bertrand than with Cournot competition.

Table 8.3 Price mark-ups

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Cournot competition						
Standard product	9 %	10 %	---	9 %	---	---
Improved product		21 %	18 %		29 %	11 %
Average	9 %	18 %	18 %	9 %	29 %	11 %
Bertrand competition						
Standard product	7 %	7 %	--	7%	--	--
Improved product		16 %	14 %		25 %	8 %
Average	7 %	14 %	14 %	7 %	25 %	8 %

Not only does the model predict that the industry will be made up of fewer firms when price rather than quantity competition is the norm, but also that they will be larger. This is true whether output levels or sales volumes are used as a proxy of the firms' size. Firm-level output is around 40% greater in each stage and for any product quality sold, while the jump in sales is in the order of 35%; the second figure is smaller because prices are lower with Bertrand competition in the product market. Equally interesting, perhaps, is that the *aggregate* sales volume is virtually the same with either price or quantity competition. In other words, greater sales at the firm-level approximately 'cancels out' with the lower equilibrium number of firms supported by the product market, and hence industry-wide sales are practically unaffected by the intensity of competition. And the same goes for the benefits to consumers. A comparison of quantity and price competition shows a modicum gain in consumer welfare of only 3 to 6% in the latter case, depending on the stage and the outcome

of the innovation process. The policy implications of these results are discussed in Section 8.2.3 together with those concerning market fragmentation.

Table 8.4 Output levels

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Cournot competition						
Traditional firm	2,273	2,381	---	2,273	---	---
Innovator		8,185	7,102		28,571	11,111
Imitator			7,102			11,111
Average	2,273	4,492	7,102	2,273	28,571	11,111
Bertrand competition						
Traditional firm	3,111	3,195	--	3,111	--	--
Innovator		11,347	9,566		37,500	15,278
Imitator			9,566			15,278
Average	3,111	6,252	9,566	3,111	37,500	15,278

Table 8.5 Firm-level profit/loss margins

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Cournot competition						
Traditional firm	0 %	1 %	---	0 %	---	---
Innovator		– 35 %	13 %		8 %	8 %
Imitator			1 %			0 %
Average	0 %	– 24 %	7 %	0 %	8 %	2 %
Bertrand competition						
Traditional firm	0 %	1 %	--	0 %	--	--
Innovator		- 25 %	11 %		9 %	6 %
Imitator			2 %			0 %
Average	0 %	- 17 %	6 %	0 %	9 %	2 %

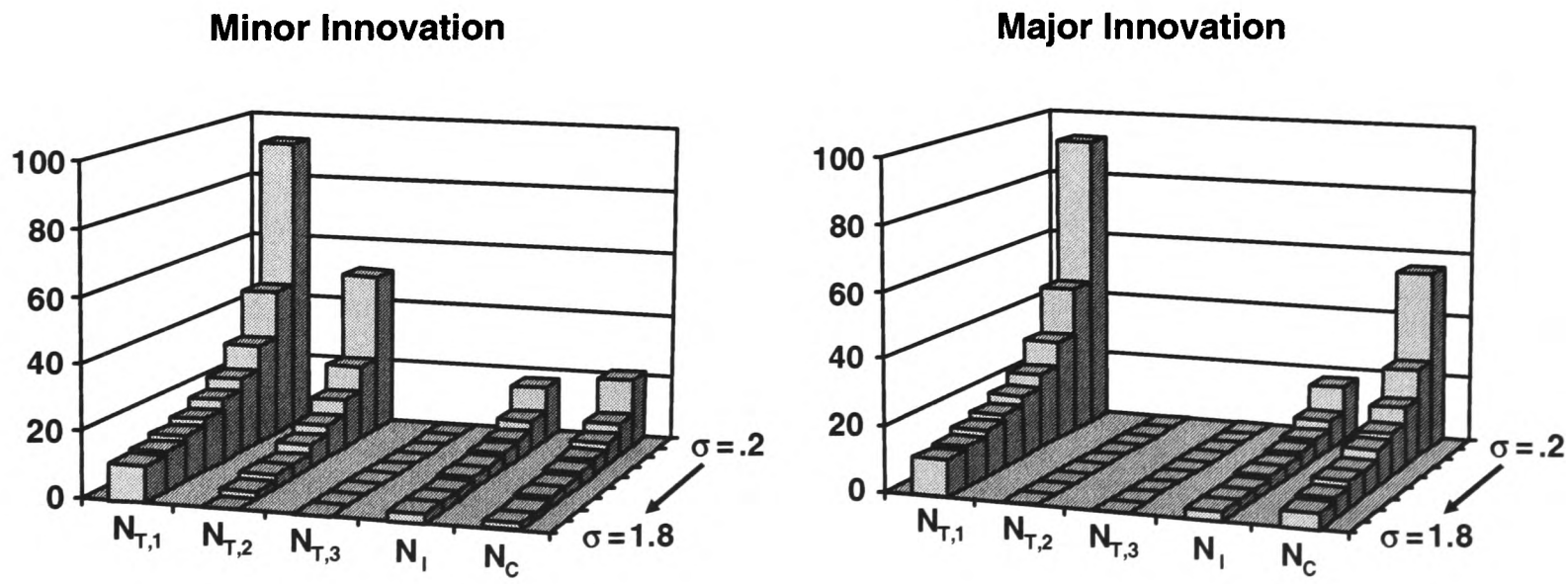
Table 8.6 Aggregate measures of profitability and welfare

	Minor Innovation			Major Innovation		
	Stage 1	**Stage 2**	**Stage 3**	**Stage 1**	**Stage 2**	**Stage 3**
Cournot competition						
Innovators' profits		– 6,970	2,183		5,927	2,069
Industrial profits	63	– 6,877	2,365	63	5,927	2,259
Consumer welfare	9,814	10,240	11,622	9,814	20,408	37,037
Social welfare	9,877	3,363	13,987	9,877	26,335	39,296
Sales volume	23,554	28,926	33,574	23,554	73,469	92,593
Bertrand competition						
Innovators' profits		-5,019	1,750		6,263	1,610
Industrial profits	52	-4,945	2,000	52	6,263	1,902
Consumer welfare	10,163	10,827	12,300	10,163	21,094	38,513
Social welfare	10,215	5,882	14,299	10,215	27,356	40,415
Sales volume	23,230	28,365	32,799	23,230	70,313	91,030

8.2.2 Market fragmentation: From niche markets to perfect substitutability

An alternative method of assessing the influence of product market features on technological change and market structure is to hold constant the mode of inter-firm competition and examine the patterns of industrial dynamics that result in markets where the degree of horizontal product differentiation (or 'fragmentation") differs. The specification of the model is well suited to this task, as the quality-augmented quadratic utility function explicitly includes a parameter measuring the inverse degree of horizontal product differentiation (σ). To pinpoint the changes in the model's solution, a comparative equilibrium dynamics analysis was carried out in which σ was varied from a value close to zero, a situation with a high degree of horizontal product differentiation, to one approaching 2, where products of the same quality are almost perfect substitutes to consumers. The complete results of this exercise are reproduced in Tables A8.1 to A8.7 in the Appendix; notice that the simulations assume that firms compete in quantities, but simulations with price competition generate similar results.

Figure 8.1 Number of firms across the sustainable equilibria



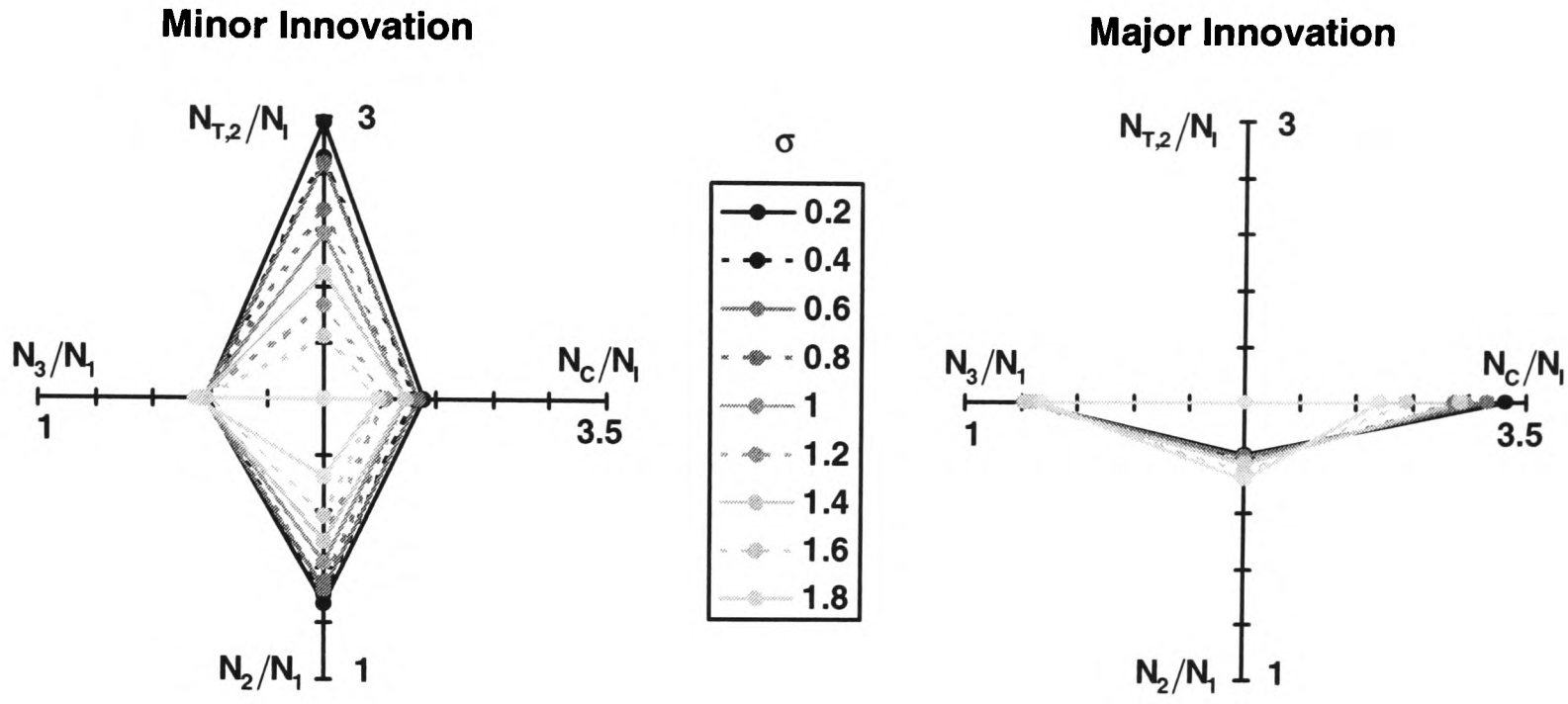
The set of results discussed in the benchmark case (σ=1) resurface in simulations run with other values of σ. The evolution in product development, market structure, pricing, firm

growth, profitability and welfare seems to be as relevant to fragmented markets with weak competition between the firms in the industry, as it is to those where the rivalry between producers is more intense because consumers discriminate between them solely on the basis of product quality. But, as in the comparison of price and quantity competition, the commonality of these general tendencies conceals a wide variation in the *scale* of industry-wide production and R&D investments supported by markets with more or less horizontal differentiation. The equilibrium number of traditional, innovative and imitative firms in each stage of the evolution is lower when horizontal differentiation is less. This is portrayed in Figure 8.1, which shows that: the number of innovating firms ranges from 17 when $\sigma = .2$, to 3 when $\sigma = 1.8$; the corresponding number of imitating firms declines from 21 to only 2 when innovators have attained a minor innovation, and drops from 55 to 5 with a major innovation; and, as for the firms producing the conventional product, while 92 are supported in the stage before innovation when $\sigma = .2$, the number falls to 11 when $\sigma = 1.8$, and the feasible entry post-innovation is lower as well.

But, at variance with the analysis of Bertrand vs. Cournot, the entry rates that can be supported do not appear to shift down 'evenly' as the substitutability between the products changes. To positively identify the 'biases' towards specific product lines and research activities as σ is varied, Figure 8.2 displays the progression in the ratios of: (i) traditional firms to innovators in stage 2, $N_{T,2}/N_I$; imitators to innovators, N_C/N_I; the total number of firms in stage 2 to the initial number of firms, N_2/N_1; and, the number of firms in stage 3 to the firms in stage 1, N_3/N_1. There one can perceive that the ratio of firms producing the conventional product line in stage 2 to those that invested in own-product innovation is lower when σ is greater. $N_{T,2}/N_I$ goes from being 3 traditional firms to each innovator when $\sigma = .2$, to zero when $\sigma = 1.8$; the results correspond to a minor innovation because the simulations indicate that (conditional on the other parameters) it is never profitable to

produce the conventional product when a major innovation has been developed, whatever the degree of horizontal differentiation. Figure 8.2 also indicates that the ratio of imitators to innovators is lower when product substitutability is greater.

Figure 8.2 Relative changes in product development and market structure



That is to say, when innovation is taking place in markets where consumers are mostly concerned with vertical quality differences among the goods supplied even a minor improvement in product design can thwart the commercial viability of the conventional product line. This is just what one would expect: when horizontal differences are less pronounced, vertical differences become much more relevant. But, in addition, the model predicts that fragmented markets will support comparatively less investment in imitative-R&D compared to innovative-R&D. All told, innovative ventures appear to be the least affected by a high degree of product substitutability. Because the formal results in Chapter 6 have shown that there is a one-to-one trade-off in the number of innovating and imitating firms in the Stage 3 equilibrium, one needs to view the varying entry ratios as a consequence

of the different pressure exerted by product substitutability on innovators' expected profits as against that of traditional producers.[4]

These considerations bring us to the question of the relationship between product differentiation and industry structure. Because there are fewer firms at each stage of the game when σ takes on a higher value, the industry would be more concentrated. The Herfindahl index in stage 1 ranges from a value of 109 when $\sigma = .2$ to 909 when $\sigma = 1.8$, almost an order of magnitude greater, and similar variations are observed for stages 2 and 3, for both outcomes of the innovation process (see Table A8.2 in the appendix).

This finding seems – at least at first blush – to agree with one of the key results of the "bounds approach" developed by John Sutton (1991, 1998). [5] Sutton shows that in industries with high R&D to sales ratios, the concentrating effect set in motion by product innovation should be positively related to the substitutability between products in the market. The reasoning behind this is simple: horizontal product differentiation protects firms from the vertical (quality) differences existing between their products. In the extreme case of complete horizontal differentiation, the profits of each firm depend only on the quality of their own product and so they would be completely insulated from innovations developed by rivals in the industry. When products are close substitutes of each other in the horizontal dimension, the process of innovation would exert a powerful influence on market structure.

[4] In particular, it invites the following conjecture: if an industry is characterised by quality heterogeneity at a given stage in its evolution, horizontal substitutability weakens the revenue position of traditional producers comparative to that of innovative firms, allowing the latter to recover more of its R&D costs early on, which then gives it an advantage compared to imitators. Though interesting, we make no attempt to prove this conjecture in the thesis.

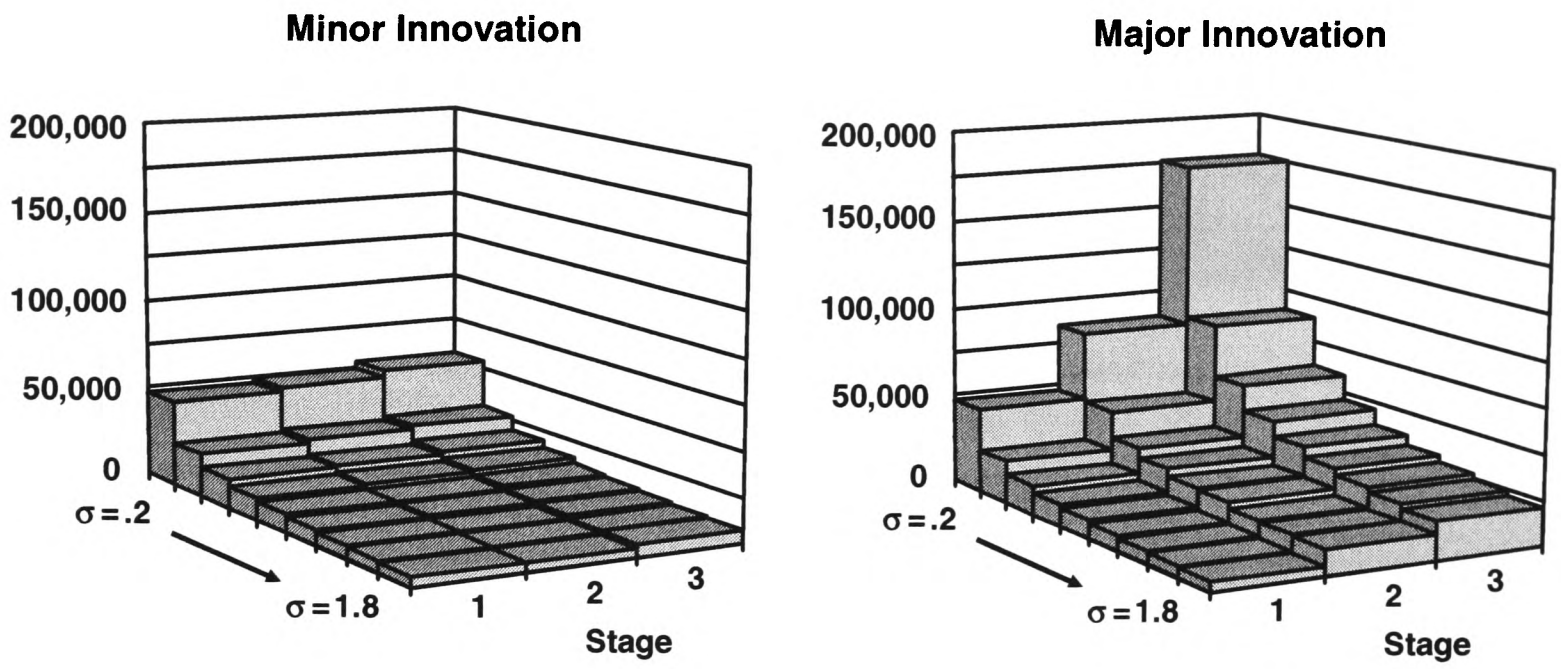
[5] The connection between our theoretical framework and John Sutton's "bounds approach" runs deeper than this. By finding the maximum number of feasible firms investing in the two classes of R&D, and the maximum number of traditional firms over the evolution of the industry, the "symmetric sustainable equilibrium" implicitly finds the (evolving) lower bound to market concentration. The results therefore determine a ceiling to the number of innovators that could emerge in an industry undergoing technological change. But one must be careful in making this analogy, as the number of traditional and imitating firms that can be supported is closely tied to the scale of innovative-R&D investments and their eventual outcome.

A closer examination of the *relative* concentration between one stage and the next in the present model, however, leads us to qualify that conclusion. Suppose the ratio of firms in stage 2 to 1 (N_2/N_1) is employed as a proxy of the concentrating effect of innovation in the interval before the new product designs can be copied – i.e., whilst innovators benefit from a "lead time"; and, correspondingly, that the ratio of firms in Stages 3 to 1 (N_3/N_1) is used as a long-run measure of relative concentration. Then, Figure 8.2, which maps the values of N_2/N_1 and N_3/N_1 as σ is varied, reveals that with a minor innovation, the short-run concentrating effect of this class of R&D appears to be stronger in fragmented markets; in the case of a major improvement, the effect is slight and in the opposite direction, with less concentration in Stage 2 in more fragmented markets. There is no connection between the long-run relative concentration and the degree of horizontal differentiation, whatever the outcome of innovation. These results show that the concentrating effects of innovative activities in the present model are mediated by the relative success of the original innovators, and can be 'neutralised' by the introduction of product clones later on.

Turning to the other firm-level variables, Tables A8.3 to A8.5 indicate that in settings where all the producers manufacture a good of identical quality in equilibrium, the mark-up on this good's price, the output of each firm and their profit margins are invariant to changes in the degree of horizontal product differentiation. This is true of the first stage, in which the entire industry is producing the conventional product line, the third stage, in which all the competing firms have procured the improved design through their investments in innovative- or imitative-R&D, and also of the second stage when the outcome of the innovation process has been a major design improvement, as production of the conventional product is then unprofitable. Conversely, in the presence of "quality heterogeneity," understood as the coexistence of firms producing goods of different quality in a given stage in the evolution of

the industry, the price mark-ups of both product lines, the output set by each firm, and the profit margins of innovators are increasing in σ.

Figure 8.3 Consumer welfare over the evolution of the industry



Lastly, we need to consider whether industry-wide sales and welfare are significantly different in markets that are more or less fragmented. With respect to aggregate sales, Table A8.6 indicates that these are negatively related to σ; in other words, increasing differentiation is synonymous with higher total sales at each stage, whatever the outcome of the innovation process. This is because (unlike the conclusions obtained from a comparison of the Bertrand and Cournot solutions) the fall in the sustainable number of firms is not counterbalanced by an increase in output levels of a sufficient magnitude. Firm-level outputs respond to the degree of horizontal differentiation exclusively when goods of different qualities are offered in equilibrium; in other cases, they are irresponsive to changes in σ. As Figure 8.3 shows, consumer welfare and σ are negatively associated as well. The fall in consumer welfare is most marked in the third stage of the evolution of an industry possessing a major product innovation, which is supportive of the contentions about: (i) the crucial role

played by imitators in delivering the benefits of successful innovation to consumers; and (ii), the relative 'bias' against imitative-R&D in industries with higher values of σ.[6]

8.2.3 Implications for competition policy

Current thinking on competition has come to recognise that greater rivalry is not always conducive to furthering the "public interest."[7] In the analysis of Bertrand and Cournot competition, and in the comparative dynamics with respect to product substitutability, we found that the benefits to consumers from more intense product market competition were almost completely wiped out by endogenous changes to the evolving industry structure.[8] Specifically, the two-pronged approach showed that added rivalry, by curbing the potential rent streams from developing or cloning improved products, constrained industry-wide spending on both classes of R&D. It also checked the entry of traditional producers, as they

[6] These differences in welfare are not entirely unexpected because σ is a taste parameter, appearing directly in the specified utility functions, independently of the interpretation it can be given as a measure of the 'toughness of competition.' An alternative reading of the comparative dynamics results is that investments in R&D are spurred on by consumers' "love of variety."

[7] As Bishop and Walker (1999, pp. 13-5) write in a recent book on the economics of EC competition law:

> "On an intuitive level, effective competition might be equated with the process of rivalry. Such a definition is intuitively appealing since rivalry is the means by which a competitively structured industry creates and confers its benefits. Moreover, competition law investigations often arise in those situations in which rivalry is eliminated, e.g., by merger or through a cartel agreement. But this definition of effective competition provides no benchmarks for how much rivalry is required for competition to be effective and invites the erroneous conclusion that the elimination of rivalry must always be deemed to be anti-competitive or restrictive of competition. ...
>
> Rather than always being anti-competitive, reductions in rivalry may be pro-competitive in the sense that they increase consumer welfare.
>
> Hence, such a definition of effective competition is entirely inappropriate for the purposes of competition law. It would make rivalry an end in itself, regardless of whether the elimination of some rivalry had any substantive negative effect on consumer welfare. One must recognise that all market economies require there to be some elimination of rivalry, since this is necessary to every integration or co-ordination of productive economic efforts and to the specialisation of effort."

[8] The comparison of the solutions with Cournot and Bertrand showed that there was only a 3 to 6 % difference in consumer welfare between the two, which seems quite small when put against the wide variations in concentration, price and output levels, and profit/loss margins. Consumer welfare was actually lower when the products sold were closer (horizontal) substitutes of each other.

need to recoup their fixed production costs. As a result, the industry would become more concentrated. The inverse relationship between the (exogenous) intensity of product market competition and the (endogenous) concentration of the industry is a consequence of the game-theoretic approach that has been postulated, which gives primacy to the sunk costs required to research, develop, and manufacture new products, while permitting an endogenous determination of the industry-wide scale of R&D investments and production. The endogeneity of the R&D investments and market structure serves as an equilibrating mechanism, which 'substitutes' the exogenous kind of product competition imposed with another in equilibrium, this being the only means by which firms can expect to break even.

Taken at face value, these results cast doubt on the desirability of public policies that aim to positively encourage more competitive forms of behaviour or to dismantle horizontal barriers (derived from, e.g., advertising expenditures) in industries where there is rapid technological change.[9] Note, in this regard, that the guidelines set by the Office of Fair Trading (OFT) in the UK make the following proviso:

> "In markets where there is a high rate of innovation it may be natural to see high prices for a period of time. Persistently high profits which result from successive innovations will provide both a return on previous innovations and incentives for further innovation. In these circumstances high profits will not indicate an abuse. ... [T]he Director General will therefore be mindful of the need not to interfere in natural market mechanisms where high prices will encourage new entry or innovation and thereby serve to increase competition."[10]

Because the solution to the model assumes a level playing field, both in resource and informational terms, we cannot give any opinion on the potential benefits of policies that do away with horizontal barriers that are asymmetric; nor on the matter of the additional costs

[9] But, as Stiglitz (1987, pp. 886-7) has argued, firms can take a number of actions that could undermine this argument: "Entry may not entail competition. The entrant or entrants and the incumbent may collude rather than compete. ... Incumbent firms will take action to deter entry; entering firms will take action to facilitate collusion."

[10] *OFT Guideline 402, The Chapter II prohibition*, sections 4.11 and 4.12. See Whish (1993), Furske (2002), or Middleton et al (2003) for a detailed examination of UK and EC competition law.

faced by new entrants if they are to enjoy the degree of horizontal differentiation of incumbents. Nevertheless, the analysis we have undertaken holds a number of more specific implications for the conduct of competition policy.

First, it leads us to question the appropriateness of relying on a static analysis of structural measures of market power to ascertain whether a firm has a "dominant position" and if it is abusing it.[11] An index such as the Herfindahl cannot properly inform policymakers of the sustainability of a monopoly or oligopoly situation because, as some in the legal profession have noted "[e]x hypothesi market shares cannot indicate the competitive pressure exerted by firms not yet operating on the market but with the capacity to enter it."[12] Take, for example, the results in Section 8.2.2. Going by the yardstick adopted by the U.S. Department of Justice, if innovators' efforts resulted in a major design improvement, serious competitive concerns would be raised in the second stage in industries with σ in excess of 0.6, as the Herfindahl index tops the 1800 mark of "very high concentration." Would this really warrant governmental action? With perfect foresight, the answer is no, as the entry of product clones in the third stage of the industry's evolution would bring about a sharp fall in concentration, well below the 1200 limit of "acceptable concentration." As competition authorities do not have perfect foresight, they will not usually be able to determine whether the concentrating effects of innovation are transitory or permanent.[13] Even so, the current doctrine on potential

[11] For a longer discussion of the difficulties of measuring market power in the "context of rapid innovation," see Ordover and Baumol (1988), Teece and Coleman (1998), and Pleatsikas and Teece (2001).

[12] Whish (1993, p. 263).

[13] And the results from the present framework imply that the ambiguity would be worsened by a lack of knowledge relating to the degree of horizontal differentiation enjoyed by the firms. Concretely, if σ is above 1.4, the short-run increase in concentration that results from a *minor* innovation also falls in the "very high concentration" bracket, but in this case there would be minimal investments in imitative-R&D, and the long-run market structure would remain highly concentrated.

competition seems too narrow, in that it does not include "new entry by someone outside the business acquiring the necessary technology and other resources."[14]

A related concern is that an exclusive reliance on structural measures may fail to give an accurate indication of the degree of competition in the relevant product market. The results presented in the preceding sections suggested that market shares exaggerate the firms' effective market power and, indeed, could misrepresent the intensity of competition. It would seem appropriate to base the decisions on price mark-ups as well. This has been the approach taken in the UK and the EC, although the criteria about what constitutes excessive pricing are different in each regulatory regime. In a case brought against *Napp Pharmaceutical Holdings Ltd*, the Director of the Office of Fair Trading in the UK argued that:

> "The prices charged by Napp for MST in the community are excessive. The Director considers that a price is excessive and an abuse if it is above that which would exist in a competitive market and where it is clear that high profits will not stimulate successful new entry within a reasonable period. Therefore, to show that prices are excessive, it must be demonstrated that (i) prices are higher than would be expected in a competitive market, and (ii) there is no effective competitor pressure to bring them down to competitive levels, nor is there likely to be."[15]

What constitutes effective competition obviously is tied up to the types of potential competition that are admitted in practice, and will be crucial to the true force of this guideline.

Now, prices may be excessive, but they may also be set temporarily below average costs for strategic purposes, referred to as predatory pricing. In this respect, one must emphasise that Cournot and Bertrand equilibrium prices exceed the marginal cost of production in the presence of horizontal and vertical product differentiation; in contrast to the traditional analysis of a duopoly or an oligopoly with a homogenous product and constant,

[14] This has been remarked on by Korah (1993) concerning the EC Commission. To broaden the definition of potential competition, it would be necessary to have better information about the length of time needed for new entrants or incumbents to market substitutes (or, as we have been calling them, clones); if this is considered to be overly long, some action would probably still be necessary.
[15] Quoted in Furske (2002, p. 248).

symmetric marginal costs, in which firms competing in prices set these at the level of marginal costs and so long as there are no fixed costs make zero profits. Yet, whether or not equilibrium prices exceed average costs depends on the outcome of the innovation process. This carries the implication that predatory pricing would be much more difficult to determine in oligopolistic industries driven by the consequences of research investments made under conditions of fundamental uncertainty. Although innovating firms do not set their prices to drive their competitors from the market in the present model, they may well price below average costs. But if they do so, it is precisely because there is effective competition from incumbent firms and potential entrants producing the conventional product, as well as entry by imitating firms. A better understanding of the relationship between technological uncertainty, pricing and profits is needed to separate the cases where there is an anti-competitive pricing policy from those where there is not.[16]

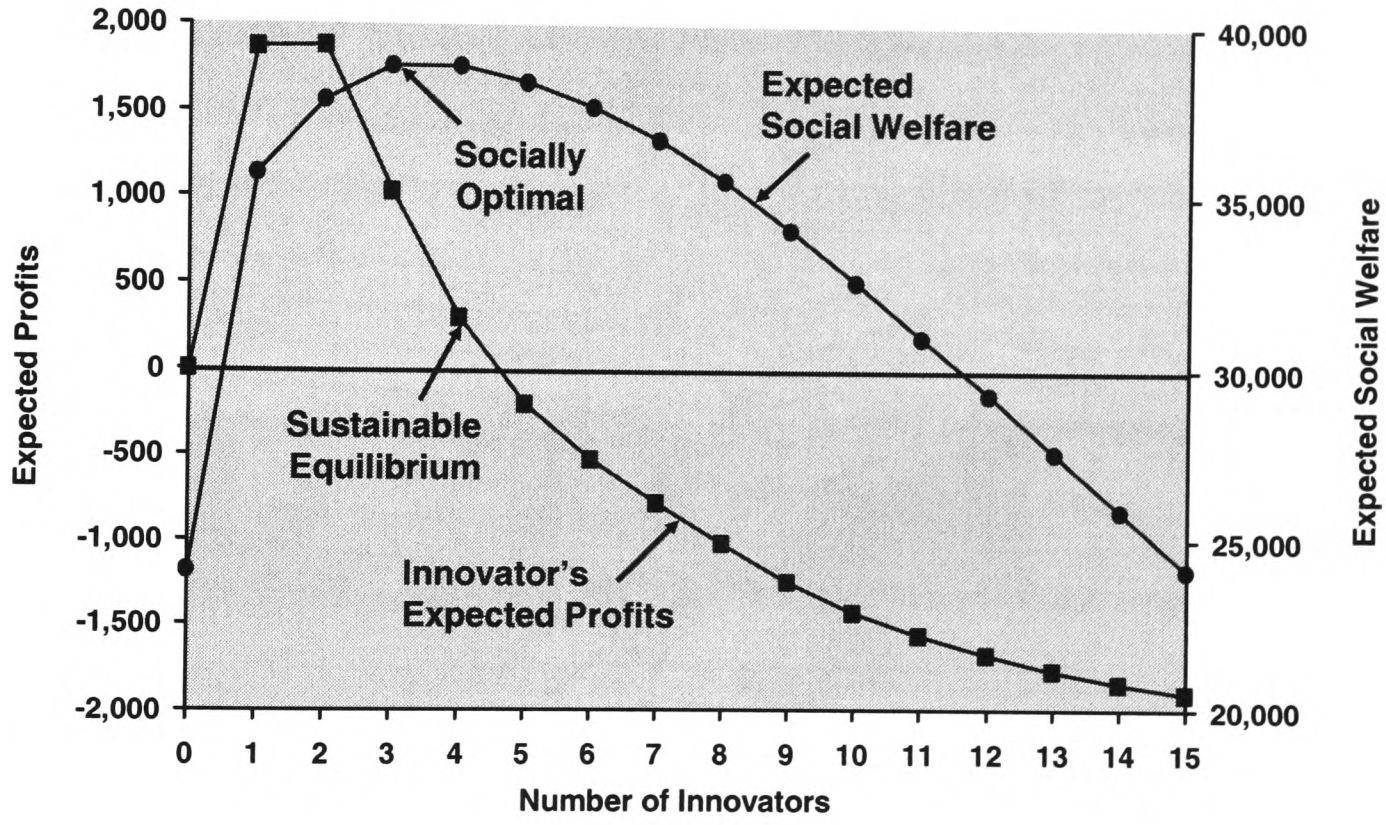
8.3 The scope for welfare-improving innovation policies

This section begins by asking whether social welfare maximisation calls for the same number of innovative firms as was derived in the sustainable equilibrium on the basis of private incentives. To answer this, Figure 8.4 overlays the expected profits of a typical innovator and expected social welfare, both as a function of the number of innovators. For purposes of comparability, expected social welfare is derived from a simulation with the benchmark parameters; we are thereby assuming that the "social welfare maximiser" has the same beliefs about the innovation process as the private sector, and that it uses the same discount rate

[16] This issue has been widely debated in legal circles. Valentine Korah, barrister and Professor of Competition Law at UCL, made the following indictment of the (paucity of) economic foundations in EC Commission decisions on competition: "It would be educational if the Commission would spell out the economic arguments for and against a particular practice more specifically. It should not merely assume that tying, compartmentalization of the market or charging below some level of cost is anti-competitive. I would like it to analyse why it is so, and under what conditions." Korah (1993, p. 188).

when adding the benefits gained by firms and consumers across the three stages. Figure 8.4 indicates that expected social welfare peaks after the maximal expected profits for a typical innovator, but before the point at which expected profits cut the zero-axis. This implies that the socially optimal number of investments in innovative-R&D is lower than the number determined by free entry conditions. But the case for governmental intervention to encourage further spending in own-product innovation is not at all strong, as the difference in expected social welfare with 3 or 4 investments is very slight indeed.

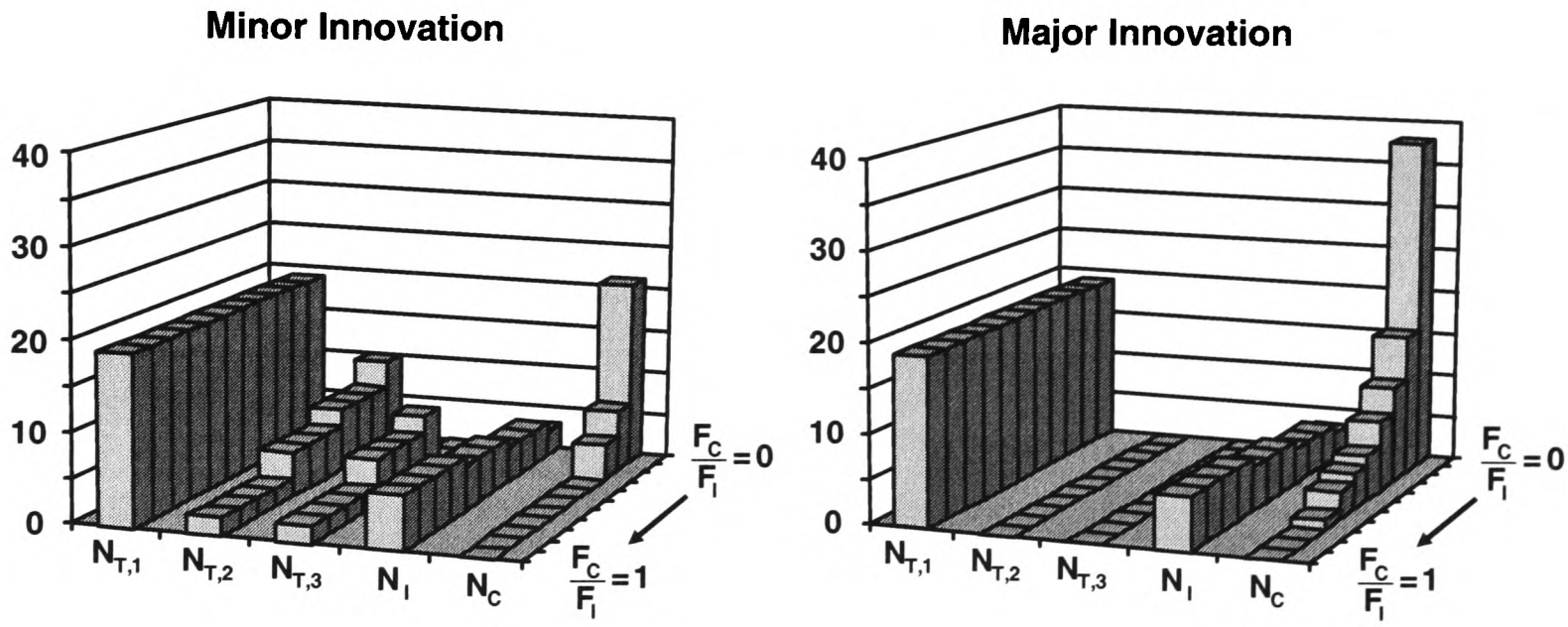
Figure 8.4 Expected profits and social welfare from innovative-R&D



Although the scope for increasing expected welfare by encouraging innovation seems to be rather limited, it may be that the case for stimulating imitative-R&D expenditures is stronger. In the model put forward, the most suitable proxy for the relative cost of original development vis-à-vis imitation is given by the ratio of the sunk costs of these activities, F_C/F_I. Accordingly, the analysis of the various instruments that the government could employ with the object of eliciting further imitation is based on a comparison of the equilibria

resulting with different values of F_C/F_I ; ranging from $F_C/F_I = 0$, in which the innovators' product designs are freely available at the third stage, to $F_C/F_I = 1$, in which the learning processes needed to copy the improved designs are as costly as the research and development efforts of the original innovators. Specifically, the comparative dynamics results in Tables A8.8-A8.14 (in the Appendix) were derived by fixing the sunk cost of innovation at $F_I = 2{,}500$ and varying F_C from 0 to 2,500, keeping all other parameters as in the benchmark case. Figure 8.5 presents the equilibrium number of traditional, innovating and imitating firms, as the ratio F_C/F_I climbs from 0 to 1.

Figure 8.5 Number of firms across the sustainable equilibria



As one may well expect, lower sunk costs of imitation result in more firms copying the new product in equilibrium: if the relative cost of imitation is sufficiently low, clones of the minor innovation are marketed; conversely, if it is high enough, not even the major innovation would be copied. Notice that, when interpreting these results, one must distinguish between the number of clones being marketed and the total amount spent on imitative-R&D, as the latter will depend on the value taken by the fixed cost F_C. Whilst the equilibrium number of imitating firms is negatively related to the sunk cost they must disburse, a calculation of the

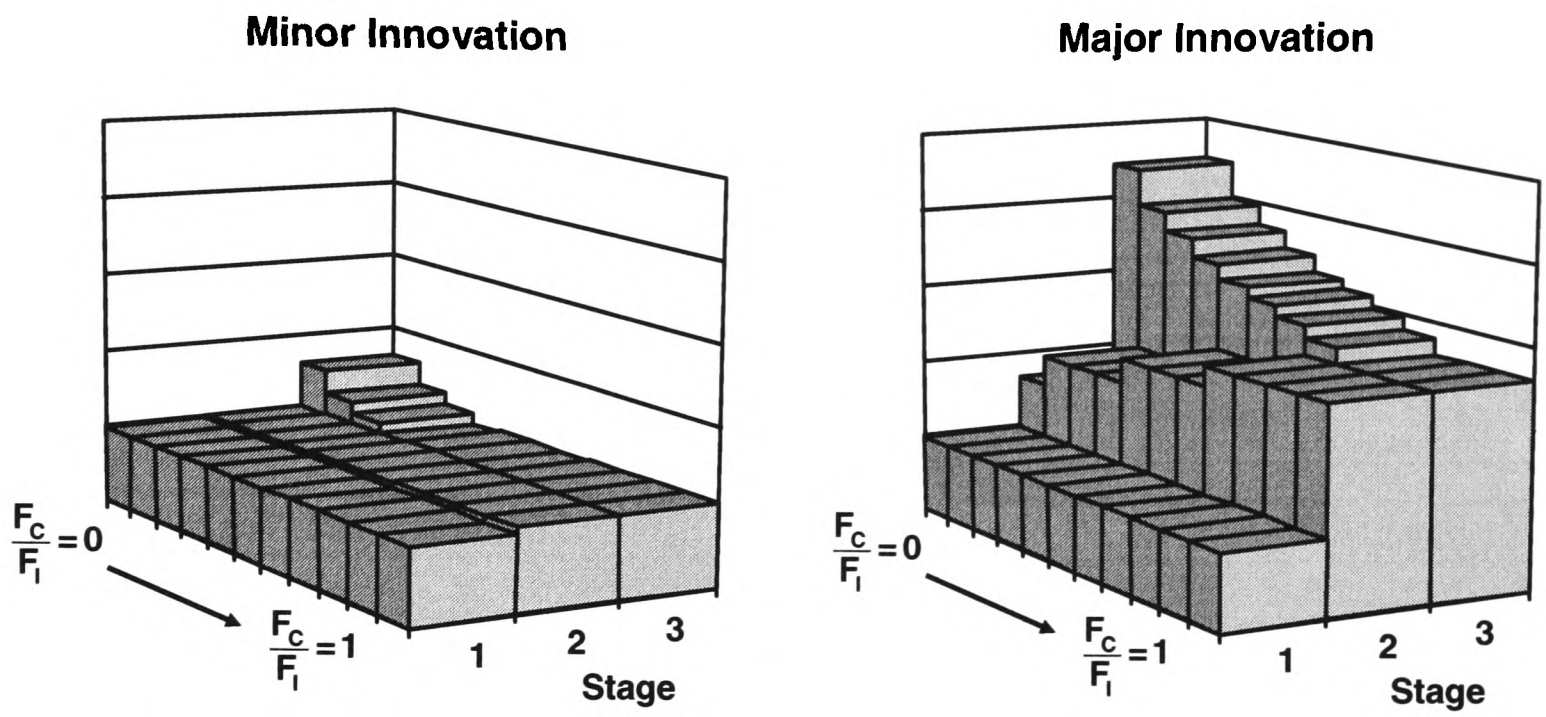
industry-wide expenditures on imitation shows this to rise as F_C/F_I increases from a value of 0, peak at $F_C/F_I = 0.3$, and then fall as F_C/F_I goes to 1. This suggests that a policy that makes imitation cheaper may get more firms to market clones and simultaneously observe a reduction in the total expenditure directed towards imitative activities.

Looking to own-product development, fewer investments are made the lower the relative cost of imitation. Given that the sunk cost of innovating F_I is the same in each simulation, this result must be viewed as a knock-on effect of the increased number of imitators that can be supported in the industry when the cost of imitation is cheaper. The heightened rivalry (and hence lower revenue) in stage 3 would be factored into the innovators' assessment of expected profits, giving rise to less industry-wide investment in innovation. As a relatively low cost of imitation 'dissuades' firms from undertaking innovative-R&D at the same time as others are 'induced' to develop product clones, traditional producers encounter a less competitive environment in the short run, but more intense rivalry in the long run. As a consequence, a greater number of firms can market the conventional product line in stage 2 when F_C/F_I is lower and, when no imitation takes place, they can continue to market this product into stage 3; when imitation does take place, the new product will displace the conventional one at the final stage. These observations relate to the evolution with a minor innovation, because (given the other parameter values) the traditional firms cannot recover their fixed production costs when facing a much-improved product.

The 'substitution' between the two classes of R&D activities as their relative cost changes has the following implications for market structure. In general, the lower is F_C/F_I, the less concentrated the industry; the one exception is the 2nd stage of an industry with a major innovation, in which the drop in innovators is not accompanied by an increase in the number of traditional firms. This pattern can be seen in Table A8.9, which also indicates that

industry concentration follows an inverted U-shaped path when the relative cost of imitative-R&D is low enough for firms to finance this research in equilibrium; the concentrating effect of innovation persists if high costs of imitation discourage the development of product clones. Similar results emerge from a consideration of price mark-ups.

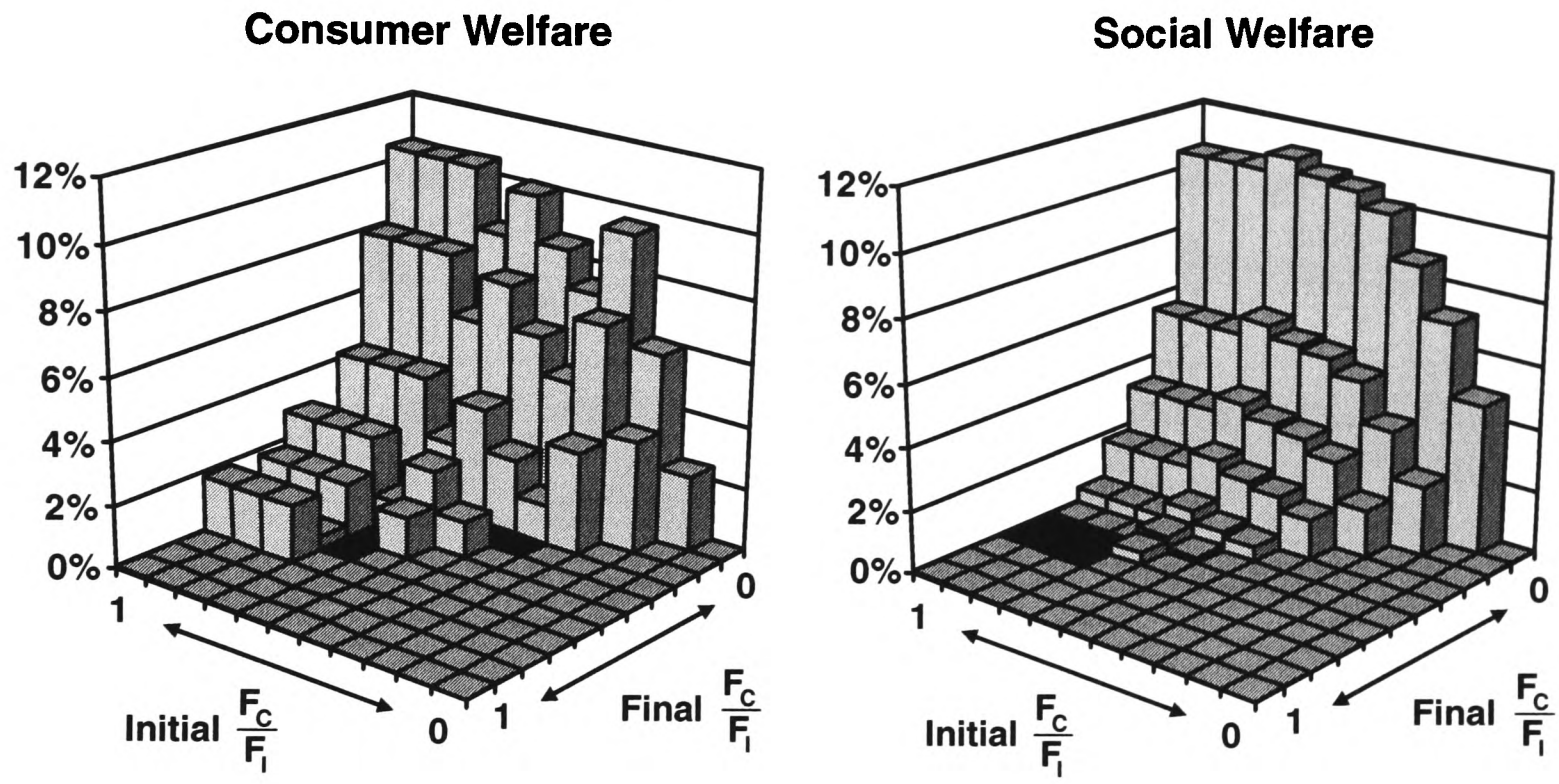
Figure 8.6 Consumer welfare over the evolution of the industry



The re-orientation of R&D spending in equilibrium results in a remarkable trade-off between industrial profits and consumer welfare across time. In policy environments facilitating imitation, the expected rewards from innovative research are undermined. This limits the wasteful duplication of high-priced R&D expenditures and makes for a less competitive market in the short run, increasing profits (limiting losses) in the industry at the cost of less consumer welfare. At the same time, by strengthening the incentives for firms to copy the improved product, such a policy will encourage a more competitive market in the long run, which cuts the rents gained by the business sector and advances consumers' interests. The trade-off can be detected by comparing Tables A8.13 and A8.14, and a graphical depiction of the evolution of consumer welfare for different values of F_C/F_I is presented in Figure 8.6 Importantly, the substitution of R&D resources that is induced by a change in their relative

costs must not be so extreme that no innovation is financed in equilibrium. If the number of firms investing in innovation is low to start with, a policy that encourages imitation may lead to a stagnant industry with no product development at all. This eventuality does not come to pass in the present simulations, but is certainly possible if the potential gains in product quality take a lower value or the sunk cost of innovation is greater.

Figure 8.7 Expected consumer and social welfare variations



A final tally of the welfare effects from changes in the relative cost of the two classes of R&D can be made on the basis of the sum of expected consumer or social welfare across time. As was noted earlier, this calculation depends on the social discount rate and the government's expectations regarding the innovation process. If the social discount rate equals the private discount rate $\delta = 0.8$, and government has the same beliefs as the private sector about the uncertainty involved in the innovation process $p = 0.5$, then the (expected, cumulative and discounted) consumer welfare is in general higher when imitation is relatively cheaper. The same is true of social welfare. As evidence for these assertions we present in Figure 8.7 a mapping of the percentage changes in expected consumer and social welfare that

would originate from variations in the ratio of sunk costs. This shows, for example, that the equilibrium industry evolution corresponding to $F_C/F_I = 0.1$ can be expected to generate approximately 4% more ('units of') consumer welfare than an industry in which the ratio of F_C to F_I is 0.2, and 2% more social welfare. The black areas in the figure correspond to a comparison for which there is a reduction in expected consumer or social welfare; thus, the expected social welfare when $F_C/F_I = 0.6$ is less than when this is $F_C/F_I = 0.8$.

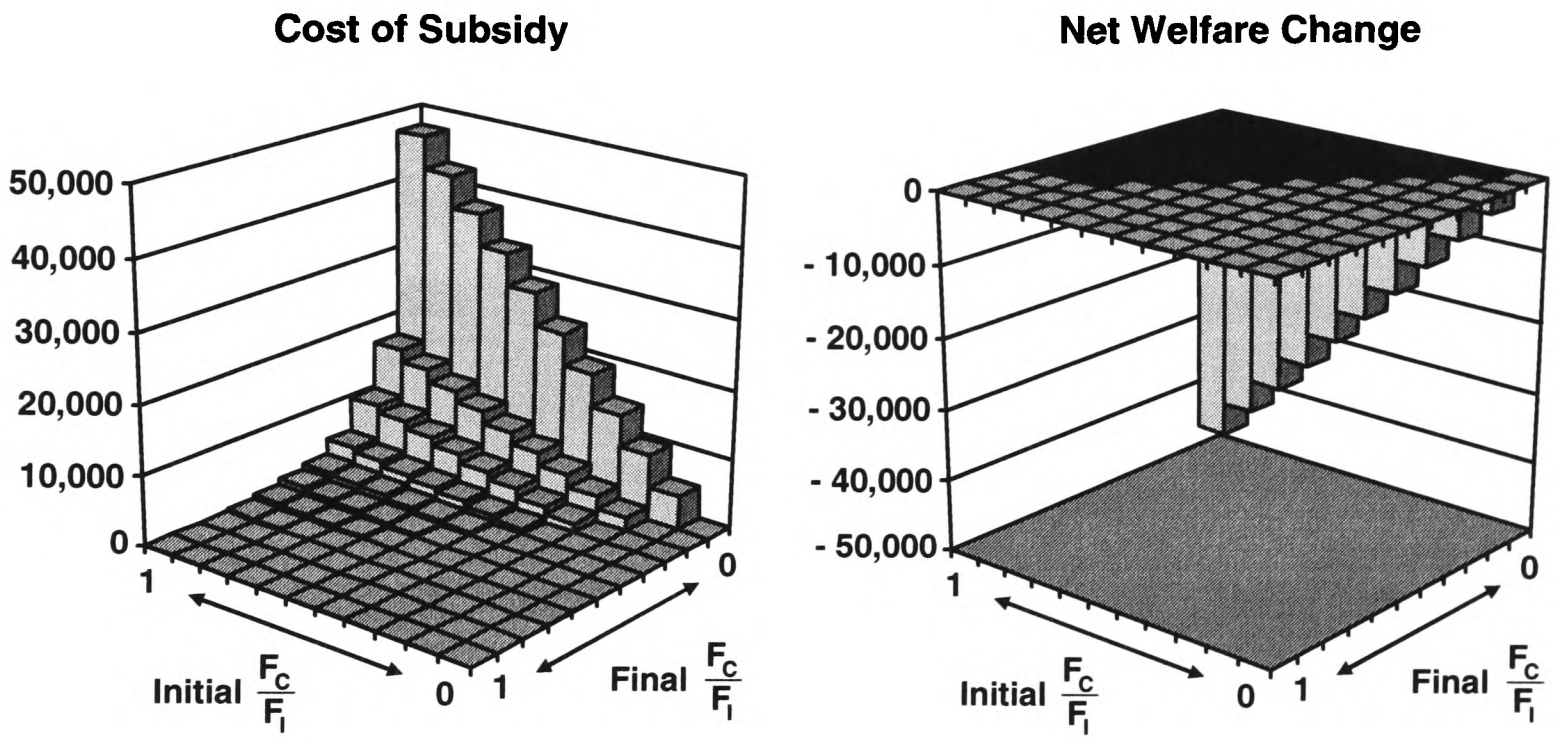
The basic implication to be drawn from this comparison is that a policy bearing down on the relative cost of imitation tends to be welfare-increasing in this modelling approach – at least, if there is no cost to the government from implementing such a policy. The final element of this inquiry has to do with the normative properties of the present analytical framework when the instrument used to 'manipulate' the relative cost of imitation has a direct cost to the government. To be precise, whether the anticipated welfare gains exceed the expected cost of a targeted subsidy given to imitating firms. Both the expected benefits and costs depend on the number of firms marketing product clones. Suppose, for example, that the government wishes to lower the ratio of F_C/F_I from 0.4 to 0.1; it follows that it should give 750 to each imitating firm in order to lower the effective sunk cost that they finance from 1,000 to 250. The expected cost to the government of this transfer can be calculated to be 5,520, which results from computing the discounted cost of the subsidy in each state of nature, and using this to obtain the expected value.[17]

The left-hand chart in Figure 8.8 plots the amount of funding that would be required to 'move' between any two combinations of F_C/F_I. Clearly, the greater the desired reduction in imitation costs paid for by the firms, the larger will be the (expected and discounted) cost

[17] The cost of the subsidy is equal to 7x750 when a minor innovation has been developed, and 16x750 with a major innovation; these values need to be discounted twice at the rate of 0.8, and added using the probability of each outcome ($p=1-p=0.5$) as a weight.

of the transfer. The cost surges upwards if the government wishes to bring F_C/F_I below 0.3, as the subsidy is then granted to a sharply increasing number of imitating firms – below this level, firms will be 'induced' to clone both the minor and major product innovations. When the cost of the subsidy is deducted from the anticipated social welfare gains and the expected net gains are plotted, the all-encompassing 'black hole' (in the right-hand side chart of Figure 8.8) indicates that it is unfeasible to use targeted subsidies to imitation to better social welfare Moreover, the results suggest that larger subsidies will result in greater social losses.[18]

Figure 8.8 The balance of welfare from promoting imitation



In summary, the normative analysis undertaken in this section suggests that social welfare maximisation calls for approximately the same level of innovative-R&D as in the symmetric sustainable equilibrium. Policies that stimulate imitative-R&D activities are estimated to be welfare-improving, but the increases in welfare are not so large as to justify the use of targeted subsidies to induce further entry of product clones. The government would need to

[18] Both in absolute terms, but also if one measures these in relative terms. If the expected welfare change for each combination is divided by the expected cost of the subsidy, the values nearest to the diagonal (which stand for the marginal changes in the relative cost of imitation) are closest to 1, the "break-even point" in welfare terms, and fall to 0 the larger the desired change in F_C to F_I.

foster the process of knowledge diffusion through policy instruments that do not involve direct transfers. These could be information provision policies, in which innovating firms are compelled to give detailed specifications of their enhanced product designs, for example.[19] Alternatively, the government could weaken the patent system. Notice, in this regard, that the comparative dynamics analysis give a better indication of the potential effects of narrowing the breadth of the patents, which would reduce the cost to imitators of reconfiguring the product innovations, than of shortening their duration, which would stimulate more rapid entry. To study the optimal patent length, one could explicitly vary the interval during which the designs of the innovating firms can be marketed without fear of entry by clones, rather than fixing it in advance. Or, to avoid changing the dynamic structure of the game, one could introduce a parameter weighting the Stage 2 revenue accruing to each innovating firm; the shorter the interval during which innovators are legally protected, the closer this parameter would be to zero. Naturally, the existence of a product lead-time effect, which is independent of any legal protections, implies that this parameter should never equal zero.

More generally, the results seem to indicate that the competitive mechanisms at play in the model are quite effective in limiting firms' market power and driving down the short- and long-run profits of innovators, to the benefit of consumers. These results are in the spirit of the "contestable markets" theory, except the recognition that there are fixed entry costs allows for a less sweeping set of conclusions; see Baumol, Panzar and Willig (1982) and Baumol (1982). It may be noted that Dasgupta and Stiglitz (1988) showed that the contestable markets theorem fails when there are even very slight increasing returns (from

[19] On the basis of a diffusion model in which potential adopters are heterogeneous, Stoneman and David (1986) concluded that, in competitive settings, "information provision policies yield definite welfare gains whereas subsidy policies may not," although "both types of policies will speed up and extend diffusion." In their model, information provision policies increases the proportion of the population that *know of* the new technology at a given point in time, whereas subsidies reduce the *cost of* acquiring the new technology. Our analysis adds force to their argument, by suggesting that subsidies to technology adopters are likely to result in a net social loss if the cost of the transfers is deducted from the increase in welfare.

fixed costs of entry), because the incumbent's threat is to make entry not worthwhile. In their own words, "[t]he presence of small entry costs can ensure that the industry is a pure monopoly; that is, such costs can render potential competition totally ineffective" (Dasgupta and Stiglitz 1988, p. 266). In a sense, the present model removes the retaliatory threat of the incumbents, by supposing that the potential entrant has a product that is unique in some respect and hence can resist predatory price-cutting by the incumbent; if not, then it would be possible for the incumbent(s) to dissuade innovators from spending anything for the first stage R&D.

To conclude, if policymakers are to improve on the balance of investments reached by unboundedly rational (albeit error making) profit-maximising agents, they must be aware of the delicate systemic linkages between anticipated and actual private benefits and the welfare effects from diverse research strategies.[20] Although individual firms may think of innovative- and imitative-R&D as alternatives, they are complementary activities from the perspective of consumers. The activities of innovating firms can, if successful, greatly benefit consumers by extending the quality range of the goods on offer, but those of imitating firms are very important also, because they bring about a reduction in prices and profit margins; both are responsible for the shakeout of lower quality varieties. The theoretical approach in Chapters 6-8 suggests that each modality of R&D has a unique role in the process of structural change: innovative-R&D is essential because it sets off the process and, to the extent that it is successful, creates openings for further investment and growth; imitative-R&D takes advantage of these openings, and impels the transition to a mature, more competitive industry; in combination, they can deliver increased performance and value to consumers.

[20] For reviews of the empirical evidence on the effects of public subsidies, and tax investment credits on R&D (and the implied welfare consequences in a partial equilibrium setting), see David, Hall and Toole (2000) and also Hall and van Reenen (2000).

8.4 Towards a policy-relevant research agenda

This dissertation has studied the dynamic allocation of R&D resources through a variety of formal models that sought to better reflect the diversity of the *objectives* of the firms participating in the innovation process, of the *research activities* they could fund to reach a specific objective, and of the *knowledge resources* gained by the firms from a particular modality of R&D. To find out whether the dynamic modelling approaches adhering to this conceptual framework produce new policy prescriptions, this chapter extended the analysis of the multistage model of innovation and imitation, which had been specified and solved in Chapters 6 and 7. The numerical results suggest that a policy fostering a more competitive atmosphere in the final product market will be welfare-improving, but that the added social benefits may not be all that large. This is because the industry structure (endogenously) becomes much more concentrated when the product market is more competitive, as firms will only finance innovative- or imitative-R&D if they expect to recover the sunk cost of this research from their future revenue stream. The fine balance struck between the equilibrium number of traditional, innovating, and imitating firms seems to be very significant to the framing of innovation policies as well: encouraging further entry by a given type of firm can have surprising consequences on the investment and entry decision of other firms, and some of these accompanying changes may not be in the interest of consumers. All told, the analysis implies that policy instruments need to exploit the synergies between different research strategies: these may be alternative from the point of view of the firms making the investments but they are complementary from that of consumers.

These policy insights, suggestive though they may seem, must be taken for what they are, the results from some initial incursions into the normative domain with a methodological

vehicle that is still very much in its prototype stage.[21] Thus far, the dynamic modelling approach has proved to be reliable, but there is much in it that could be elaborated and improved. The specific assumptions having to do with the cost and demand structure, the potential range of outcomes from research directed towards innovation and imitation, the nature of the firms' beliefs regarding technological uncertainty, and the sequence of entry decisions were described in the last section of Chapter 6, which also suggested how these assumptions could be relaxed in further work. The same goes for the sequence of models presented in Chapters 3 to 5. They too made a variety of assumptions about the functional forms, the interrelatedness of different variables, etcetera; and the conclusions to each of those chapters explained which assumptions were more essential to the solution, and how each limited the formal and numerical results. Because these extensions have been discussed at some length in those sections, no more will be written about them here. Together, they constitute (what could be called) a theory-oriented agenda for future research, in that it is concerned primarily with making the conditioning clauses attached to the results less restrictive. In the process, the scope of application of the model may be enlarged in policy-relevant directions, but this will be an unintended bonus rather than the premeditated result from the research effort.

Instead, the remainder of this section sets out a policy-relevant research agenda. That is, one in which the research extensions deliberately aim to bring the models into closer conformity with the various 'milieus' within which firms operate in different parts of the world today. The heuristic distinction between research that is mainly occupied with "scientific inquiries" and that having as its aim to examine "practical issues" was very aptly

[21] One may do well in recalling the warning that Hayek (1945, p. 530) once made of the applicability of equilibrium analysis: "when it comes to the point where it misleads some of our leading thinkers into believing that the situation which it describes has direct relevance to the solution of practical problems, it is time that we remember that it does not deal with the social process at all and that it is no more than a useful preliminary to the study of the main problem."

put by Alfred Marshall in Chapter III of the *Principles of Economics*, where he deals with the proper use and potential abuse of "Economic Generalizations or Laws":

> "Some parts of economics are relatively abstract or pure, because they are concerned mainly with broad general propositions: for, in order that a proposition may be of broad application it must necessarily contain few details: it cannot adapt itself to particular cases; and if it points to any prediction, that must be governed by a strong conditioning clause in which a large meaning is given to the phrase "other things being equal." Other parts are relatively applied, because they deal with narrower questions more in detail; they take account of local and temporary elements; and they consider economic conditions in fuller and closer relation to other conditions of life." (Marshall 1956, p. 31).

What are the local and temporary elements, the economic conditions that seem to be most relevant to the dynamic processes of R&D allocation at the firm- and industry-level we have been modelling? Our reading of the empirical industrial organization literature on R&D (in Chapter 2) suggests that it is the distinctive features of the labour market for research workers and the supply of finance for innovative ventures that need to be explicitly analysed if the models are to be truly relevant from a policy perspective.

The complexities of these two areas were assumed away in the previous chapters. In respect of the labour market for research workers, Chapters 3 to 5 assumed that there was a perfectly elastic supply of researchers, allowing the monopolist to hire any number of them at the going market wage. The firm had no doubt as to their capacity or trustworthiness, and therefore their marginal productivity could be precisely calculated. Similarly, Chapters 6 to 8 assumed that the sunk costs of both kinds of R&D were irresponsive to the demand by the firm or firms in question, which implicitly calls for a perfectly elastic supply of research workers. Their productivity, however, depended on the modality of R&D in which they were engaged. Since the outcome of imitative-R&D was certain, the productivity of the researchers employed in this task was known in advance. But innovative-R&D was taken to be fundamentally uncertain; the likely outcome, and hence the productivity of the researchers employed, could not be estimated exactly beforehand and could end up being much higher or

lower than anticipated. But, evidently, the uncertainty about the researchers' productivity depended on the activity they were asked to carry out, not on their own capacities, which were also taken to be homogenous.[22] In relation to finance, each of the models assumed that there was a well-functioning capital market, which allowed incumbent firms and new entrants to borrow any amount at a specified interest rate. As we remarked at the end of Chapter 6, this carried the implication that financiers had the same beliefs as the firms about the nature of uncertainty surrounding technological advances from innovative-R&D.

This institutional setting was selected, not for its parallels to any particular industrial context, but because it allowed the model-building efforts to focus on the intricacies pertaining to the purposive, privately-sponsored accumulation of valuable knowledge. One can, nonetheless, think of some industries undergoing quite rapid technological change where the firms, especially the incumbent firms, enjoy the freedom of action granted by a virtually unrestricted supply of well-trained and highly-motivated researchers and good access to financial resources to fund their uncertain research activities. Pharmaceutical conglomerates (which we commented on briefly in Chapter 6) or larger firms active in the computer hardware industry may not have any resource constraints worth modelling. In those settings, the policy advice offered in this chapter would be quite appropriate.

However, for the great majority of firms in 'new industries' the situation is one of rather acute constraints, although the nature of the constraints differs from one industry to another and from one country to the next. The variations in the 'institutional constellation' facing firms in different locations have been given some order by a recent strand of the institutional literature that has featured the contributions of Amable et al (1997), Amable and

[22] Otherwise, it would be even harder to justify the symmetry of the outcomes across firms: one would need to assume that each firm investing in R&D employed a group of workers that had the same 'mix' of talents. At any rate, we should stress that that assumption was made purely for reasons of tractability, and one would unquestionably want to relax it in later work.

Petit (2001) and Amable (2002). These suggest that the observed innovative dynamism and aggregate growth performance of different economies have been closely tied to the 'institutional constellation' in those countries. In particular, the countries that have taken the lead in technologically progressive industries share the so-called 'market-based' or 'Anglo-Saxon' system. Countries in this group, which include the U.S. and the U.K., have an institutional infrastructure comprising of: a flexible labour market, in which hiring and firing of employees is relatively simple and unions are not as prominent in the wage-bargaining process; a sophisticated financial system with good access to venture capital for start-ups and an active market for corporate control; strong protection of intellectual property rights; less governmental regulation of the product markets; and an educational system that is more oriented towards providing general skills. This is in contrast to the array of institutions in Continental European countries, in Japan or in social democratic countries of Scandinavia.

If countries can be grouped in terms of their 'institutional infrastructure,' then a policy-relevant research agenda should ideally be structured to provide models that bear a close correspondence to the *sets of institutions* that are generally observed. The four lines of research presented below have exactly this goal in mind. Note that the proposals are couched in terms of the game theoretic model of innovation and imitation, but they could be rearticulated to extend the dynamic model of exploratory- and applications-oriented research along the same axes. The first line of research would investigate the consequences for the financing of R&D and the evolution of industrial structure from having a less flexible market for R&D workers. Given the sequential structure of the approach put forward, an imperfectly competitive labour market may be thought to give firms with established product lines an advantage over entrants wishing to hire researchers; in the same way, the companies that financed own-product innovation early on may have some monopsony power over their workers, which could obstruct the entry by companies trying to clone the improved products.

Even if these labour market rigidities do not create outright barriers to entry, an inelastic supply of labour may place serious constraints on firms' growth, which would depress the potential stream of profits accruing from innovative or imitative efforts. The form taken by the rigidities can thus be expected to generate quite different outcomes on the magnitude and direction of industry-wide R&D and the dynamic patterns of output and pricing.

The second research line would study the role of financial market characteristics in industrial dynamics. If there is a well-functioning capital market, new firms can readily be established by entrepreneurs to take advantage of any investment opportunities they perceive. But in less developed financial markets, the financing options open to firms without a proven track record in production or research may be costly or nonexistent. One of the objectives of this extension would be to examine the effects of specific financing shortages on the admissible range of equilibrium dynamic sequences. The second goal would be to study the influence of differential beliefs held by creditors and firms about the likelihood of innovating. A divergence is expected because investors have less information than the firms' managers and scientists about the prospects of innovating, and are generally more averse to risk and uncertainty. This line of analysis may point to biases in the allocation of R&D resulting from the different attitudes towards fundamental uncertainty of the participants in the technological innovation process, as opposed to financial imperfections. In addition, one would want to specify the contractual arrangements between innovating firms and creditors, and between principals and workers in the firm, with much more care. These arrangements are important because they determine the eventual distribution of the profits or losses of the innovating firms and, as a result, can improve or worsen the long-term prospects of survival of uncertainty-bearing firms.

The third area of scrutiny concerns innovation policies designed to alter the relative costs of original development vis-à-vis imitation. This chapter made some progress in this

area, by looking at the potential for the government to tip the balance between innovative and imitative efforts to increase the expected social welfare, and considering whether such interventions are worthwhile if the government is using a costly instrument to influence the firms' incentives. As we remarked, these investigations are suggestive of the normative properties of the present analytical framework, but need to be supplemented by a more rigorous analysis. With the objective of making a stronger case for the strengthening or weakening of the ownership rights of the innovating firms, this study would benefit from a refined rendering of the intellectual property rights (IPR) regime and its relation to the sunk costs of R&D. More details need to be given about the length and breadth of the patents held by innovating firms vis-à-vis the lead time they may have, which itself would depend on whether they rely on patenting or secrecy to protect their product innovations. A more precise description of the meaning given to an IPR regime within the model is also essential to understand the effects on the equilibrium intensities of innovation and imitation produced by amending this legal institution as against the government's use of economic mechanisms – for instance, subsidies to a particular kind of investment.

The results of the analyses into the most salient institutions impinging on the process of technological change could be pooled together with the examination of product market features presented in Section 8.2 of this chapter to ascertain whether the present framework is consistent with the empirical observations of the institutional literature. The first element of this assessment would be to determine the range of equilibrium industrial dynamics that are consistent with the 'institutional constellation' identified in each social system of innovation and production, and to compare this to the observed industrial trends of the countries seen to belong to each system. By drawing on the results from this enquiry, the analysis could try to establish whether any theoretical underpinnings may be provided for the idea that certain institutional constellations are more favourable to innovation, diffusion and growth. In this

respect, one would also want to test whether any complementarities exist among institutions in different areas of the economy. This is one of the most compelling ideas put forward by this literature, because it implies that reforming one institution may fail to generate the desired innovative dynamism if the institutional infrastructure is 'inappropriate' in other areas. This could be accomplished by estimating how the marginal benefits to having a particular institution, say a more flexible labour market, depend on the existence of a set of other institutions. Given that precise measures of social welfare can be generated in the present approach, one could readily evaluate the aggregate benefits derived from different institutional configurations and compare them to make this assessment.

Appendix. Comparative equilibrium dynamics results

With respect to the degree of horizontal differentiation (σ)

Table A8.1 Number of firms in 'sustainable equilibrium'

σ	Minor Innovation					Major Innovation				
	Traditional firms			Innovative firms	Imitating firms	Traditional firms			Innovative firms	Imitating firms
	Stage 1	Stage 2	Stage 3			Stage 1	Stage 2	Stage 3		
0.2	92	50	0	17	21	92	0	0	17	55
0.4	46	23	0	9	10	46	0	0	9	27
0.6	31	15	0	6	7	31	0	0	6	18
0.8	23	10	0	5	5	23	0	0	5	13
1	19	7	0	4	4	19	0	0	4	11
1.2	16	4	0	4	3	16	0	0	4	8
1.4	14	4	0	3	3	14	0	0	3	8
1.6	12	2	0	3	2	12	0	0	3	6
1.8	11	0	0	3	2	11	0	0	3	5

Table A8.2 Indicators of market structure

σ	Minor Innovation						Major Innovation					
	Total number of firms			Herfindahl-Hirschman Index			Total number of firms			Herfindahl-Hirschman Index		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0.2	92	67	38	109	201	263	92	17	72	109	588	139
0.4	46	32	19	217	426	526	46	9	36	217	1,111	278
0.6	31	21	13	323	658	769	31	6	24	323	1,667	417
0.8	23	15	10	435	931	1,000	23	5	18	435	2,000	556
1	19	11	8	526	1,260	1,250	19	4	15	526	2,500	667
1.2	16	8	7	625	1,659	1,429	16	4	12	625	2,500	833
1.4	14	7	6	714	1,986	1,667	14	3	11	714	3,333	909
1.6	12	5	5	833	2,546	2,000	12	3	9	833	3,333	1,111
1.8	11	3	5	909	3,333	2,000	11	3	8	909	3,333	1,250

Table A8.3 Price mark-ups

σ	Minor Innovation					Major Innovation				
	Standard product			Improved product		Standard product			Improved product	
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3
0.2	9 %	9 %	---	18 %	18 %	9 %	---	---	28 %	11 %
0.4	9 %	9 %	---	18 %	18 %	9 %	---	---	28 %	11 %
0.6	9 %	9 %	---	19 %	18 %	9 %	---	---	29 %	11 %
0.8	9 %	9 %	---	20 %	18 %	9 %	---	---	28 %	11 %
1	9 %	10 %	---	21 %	18 %	9 %	---	---	29 %	11 %
1.2	9 %	9 %	---	22 %	18 %	9 %	---	---	26 %	12 %
1.4	9 %	10 %	---	23 %	18 %	9 %	---	---	29 %	11 %
1.6	9 %	10 %	---	24 %	19 %	9 %	---	---	28 %	12 %
1.8	9 %	---	---	26 %	18 %	9 %	---	---	26 %	12 %

Table A8.4 Output levels

σ	Minor Innovation					Major Innovation				
	Traditional firm			Improved product		Standard product			Improved product	
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3
0.2	2,252	2,257	---	6,933	6,853	2,252	---	---	27,778	10,989
0.4	2,273	2,287	---	7,199	6,975	2,273	---	---	27,778	11,111
0.6	2,273	2,298	---	7,468	6,975	2,273	---	---	28,571	11,236
0.8	2,315	2,262	---	7,710	6,975	2,315	---	---	27,778	11,364
1	2,273	2,381	---	8,185	7,102	2,273	---	---	28,571	11,111
1.2	2,273	2,304	---	8,461	6,975	2,273	---	---	26,316	11,628
1.4	2,252	2,404	---	9,014	7,102	2,252	---	---	29,412	11,111
1.6	2,315	2,404	---	9,515	7,512	2,315	---	---	27,778	11,905
1.8	2,273	---	---	10,280	6,975	2,273	---	---	26,316	12,048

Table A8.5 Firm-level profit/loss margins

σ	Minor Innovation						Major Innovation					
	Traditional firms			Innovative firms		Imitating firms	Traditional firms			Innovative firms		Imitating firms
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3		Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	
0.2	0 %	0 %	---	– 49 %	12 %	0 %	0 %	---	---	7 %	8 %	0 %
0.4	0 %	0 %	---	– 45 %	13 %	1 %	0 %	---	---	7 %	8 %	0 %
0.6	0 %	0 %	---	– 42 %	13 %	1 %	0 %	---	---	8 %	9 %	0 %
0.8	1 %	0 %	---	– 40 %	13 %	1 %	1 %	---	---	7 %	9 %	1 %
1	0 %	1 %	---	– 35 %	13 %	1 %	0 %	---	---	8 %	8 %	0 %
1.2	0 %	0 %	---	– 33 %	13 %	1 %	0 %	---	---	5 %	9 %	1 %
1.4	0 %	1 %	---	– 28 %	13 %	1 %	0 %	---	---	9 %	8 %	0 %
1.6	1 %	1 %	---	– 24 %	14 %	3 %	1 %	---	---	7 %	9 %	2 %
1.8	0 %	---	---	– 19 %	13 %	1 %	0 %	---	---	5 %	9 %	2 %

Table A8.6 Industry-wide sales and profitability

σ	Minor Innovation						Major Innovation					
	Sales Volume			Industrial Profits			Sales Volume			Industrial Profits		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0.2	112,937	130,898	153,052	134	– 33,649	8,544	112,937	301,698	439,077	134	21,386	8,773
0.4	57,025	67,064	78,100	152	– 17,325	4,933	57,025	159,722	222,222	152	11,322	5,122
0.6	38,430	45,504	53,437	102	– 11,233	3,297	38,430	110,204	149,981	102	8,890	3,750
0.8	29,085	35,409	41,105	165	– 9,173	2,728	29,085	88,735	113,895	165	6,290	3,322
1	23,554	28,926	33,574	63	– 6,877	2,365	23,554	73,469	92,593	63	5,927	2,259
1.2	19,835	25,619	28,774	53	– 6,710	2,160	19,835	66,482	77,880	53	3,450	2,912
1.4	17,186	21,912	25,181	20	– 4,617	1,774	17,186	57,093	67,901	20	5,176	1,690
1.6	15,175	20,385	22,392	86	– 4,292	2,112	15,175	53,241	59,949	86	3,774	2,478
1.8	13,636	19,477	20,553	36	– 3,742	1,614	13,636	49,861	53,999	36	2,588	2,506

Table A8.7 Consumer and social welfare

σ	Minor Innovation						Major Innovation					
	Consumer Welfare			Social Welfare			Consumer Welfare			Social Welfare		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0.2	47,135	49,898	53,682	47,269	16,249	62,226	47,135	85,262	176,066	47,269	106,649	184,839
0.4	23,760	25,157	27,217	23,912	7,832	32,150	23,760	45,139	88,889	23,912	56,461	94,011
0.6	16,012	16,885	18,622	16,115	5,652	21,918	16,012	30,612	59,841	16,115	39,502	63,591
0.8	12,078	12,877	14,325	12,243	3,704	17,053	12,078	25,077	45,325	12,243	31,367	48,647
1	9,814	10,240	11,622	9,877	3,363	13,987	9,814	20,408	37,037	9,877	26,335	39,296
1.2	8,264	8,720	10,027	8,317	2,010	12,187	8,264	19,391	30,827	8,317	22,841	33,740
1.4	7,173	7,373	8,716	7,193	2,756	10,490	7,173	15,571	27,160	7,193	20,747	28,851
1.6	6,301	6,484	7,584	6,387	2,192	9,696	6,301	15,046	23,597	6,387	18,820	26,074
1.8	5,682	5,681	7,162	5,718	1,939	8,776	5,682	14,543	21,193	5,718	17,131	23,700

With respect to the relative R&D sunk costs (F_C/F_I)

Table A8.8 Number of firms in 'sustainable equilibrium'

F_C/F_I	Minor Innovation					Major Innovation				
	Traditional firms			Innovative firms	Imitating firms	Traditional firms			Innovative firms	Imitating firms
	Stage 1	Stage 2	Stage 3			Stage 1	Stage 2	Stage 3		
0	19	10	0	3	21	19	0	0	3	38
0.1	19	7	0	4	7	19	0	0	4	16
0.2	19	7	0	4	4	19	0	0	4	11
0.3	19	7	7	4	0	19	0	0	4	8
0.4	19	5	5	5	0	19	0	0	5	5
0.5	19	5	5	5	0	19	0	0	5	4
0.6	19	5	5	5	0	19	0	0	5	3
0.7	19	2	2	6	0	19	0	0	6	1
0.8	19	2	2	6	0	19	0	0	6	0
0.9	19	2	2	6	0	19	0	0	6	0
1	19	2	2	6	0	19	0	0	6	0

Table A8.9 Indicators of market structure

F_C / F_I	Minor Innovation						Major Innovation					
	Total number of firms			Herfindahl-Hirschman Index			Total number of firms			Herfindahl-Hirschman Index		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0	19	13	24	526	1,108	417	19	3	41	526	3,333	244
0.1	19	11	11	526	1,260	909	19	4	20	526	2,500	500
0.2	19	11	8	526	1,260	1,250	19	4	15	526	2,500	667
0.3	19	11	11	526	1,260	1,260	19	4	12	526	2,500	833
0.4	19	10	10	526	1,316	1,316	19	5	10	526	2,000	1,000
0.5	19	10	10	526	1,316	1,316	19	5	9	526	2,000	1,111
0.6	19	10	10	526	1,316	1,316	19	5	8	526	2,000	1,250
0.7	19	8	8	526	1,429	1,429	19	6	7	526	1,667	1,429
0.8	19	8	8	526	1,429	1,429	19	6	6	526	1,667	1,667
0.9	19	8	8	526	1,429	1,429	19	6	6	526	1,667	1,667
1	19	8	8	526	1,429	1,429	19	6	6	526	1,667	1,667

Table A8.10 Price mark-ups

F_C / F_I	Minor Innovation					Major Innovation				
	Standard product			Improved product		Standard product			Improved product	
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3
0	9 %	9 %	---	21 %	7 %	9 %	---	---	33 %	5 %
0.1	9 %	10 %	---	21 %	14 %	9 %	---	---	29 %	9 %
0.2	9 %	10 %	---	21 %	18 %	9 %	---	---	29 %	11 %
0.3	9 %	10 %	10 %	21 %	21 %	9 %	---	---	29 %	13 %
0.4	9 %	9 %	9 %	21 %	21 %	9 %	---	---	25 %	15 %
0.5	9 %	9 %	9 %	21 %	21 %	9 %	---	---	25 %	17 %
0.6	9 %	9 %	9 %	21 %	21 %	9 %	---	---	25 %	18 %
0.7	9 %	9 %	9 %	21 %	21 %	9 %	---	---	22 %	20 %
0.8	9 %	9 %	9 %	21 %	21 %	9 %	---	---	22 %	22 %
0.9	9 %	9 %	9 %	21 %	21 %	9 %	---	---	22 %	22 %
1	9 %	9 %	9 %	21 %	21 %	9 %	---	---	22 %	22 %

Table A8.11 Output levels

F_C / F_I	Minor Innovation					Major Innovation				
	Traditional firm			Improved product		Standard product			Improved product	
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3
0	2,273	2,344	---	8,138	2,894	2,273	---	---	33,333	4,545
0.1	2,273	2,381	---	8,185	5,580	2,273	---	---	28,571	8,696
0.2	2,273	2,381	---	8,185	7,102	2,273	---	---	28,571	11,111
0.3	2,273	2,381	2,381	8,185	8,185	2,273	---	---	28,571	13,333
0.4	2,273	2,244	2,244	8,013	8,013	2,273	---	---	25,000	15,385
0.5	2,273	2,244	2,244	8,013	8,013	2,273	---	---	25,000	16,667
0.6	2,273	2,244	2,244	8,013	8,013	2,273	---	---	25,000	18,182
0.7	2,273	2,273	2,273	8,049	8,049	2,273	---	---	22,222	20,000
0.8	2,273	2,273	2,273	8,049	8,049	2,273	---	---	22,222	22,222
0.9	2,273	2,273	2,273	8,049	8,049	2,273	---	---	22,222	22,222
1	2,273	2,273	2,273	8,049	8,049	2,273	---	---	22,222	22,222

Table A8.12 Firm-level profit/loss margins

F_C/F_I	Minor Innovation						Major Innovation					
	Traditional firms			Innovative firms		Imitating firms	Traditional firms			Innovative firms		Imitating firms
	Stage 1	Stage 2	Stage 3	Stage 2	Stage 3		Stage 1	Stage 2	Stage 3	Stage 2	Stage 3	
0	0 %	1 %	---	– 36 %	0 %	0 %	0 %	---	---	13 %	0 %	0 %
0.1	0 %	1 %	---	– 35 %	9 %	2 %	0 %	---	---	8 %	6 %	1 %
0.2	0 %	1 %	---	– 35 %	13 %	1 %	0 %	---	---	8 %	8 %	0 %
0.3	0 %	1 %	1 %	– 35 %	15 %	---	0 %	---	---	8 %	10 %	1 %
0.4	0 %	0 %	0 %	– 37 %	15 %	---	0 %	---	---	3 %	12 %	1 %
0.5	0 %	0 %	0 %	– 37 %	15 %	---	0 %	---	---	3 %	13 %	0 %
0.6	0 %	0 %	0 %	– 37 %	15 %	---	0 %	---	---	3 %	14 %	0 %
0.7	0 %	0 %	0 %	– 36 %	15 %	---	0 %	---	---	– 1 %	16 %	1 %
0.8	0 %	0 %	0 %	– 36 %	15 %	---	0 %	---	---	– 1 %	17 %	---
0.9	0 %	0 %	0 %	– 36 %	15 %	---	0 %	---	---	– 1 %	17 %	---
1	0 %	0 %	0 %	– 36 %	15 %	---	0 %	---	---	– 1 %	17 %	---

Table A8.13 Industry-wide sales and profitability

F_C/F_I	Minor Innovation						Major Innovation					
	Sales Volume			Industrial Profits			Sales Volume			Industrial Profits		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0	23,554	27,568	37,294	63	– 5,158	172	23,554	66,667	97,417	63	8,867	136
0.1	23,554	28,926	35,077	63	– 6,877	1,535	23,554	73,469	94,518	63	5,927	1,561
0.2	23,554	28,926	33,574	63	– 6,877	2,365	23,554	73,469	92,593	63	5,927	2,259
0.3	23,554	28,926	28,926	63	– 6,877	3,123	23,554	73,469	90,667	63	5,927	3,467
0.4	23,554	30,254	30,254	63	– 8,887	3,613	23,554	78,125	88,757	63	2,625	5,834
0.5	23,554	30,254	30,254	63	– 8,887	3,613	23,554	78,125	87,500	63	2,625	6,600
0.6	23,554	30,254	30,254	63	– 8,887	3,613	23,554	78,125	85,950	63	2,625	7,923
0.7	23,554	31,603	31,603	63	– 10,617	4,383	23,554	81,481	84,000	63	– 785	11,550
0.8	23,554	31,603	31,603	63	– 10,617	4,383	23,554	81,481	81,481	63	– 785	14,215
0.9	23,554	31,603	31,603	63	– 10,617	4,383	23,554	81,481	81,481	63	– 785	14,215
1	23,554	31,603	31,603	63	– 10,617	4,383	23,554	81,481	81,481	63	– 785	14,215

Table A8.14 Consumer and social welfare

F_C/F_I	Minor Innovation						Major Innovation					
	Consumer Welfare			Social Welfare			Consumer Welfare			Social Welfare		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
0	9,814	10,142	16,075	9,877	4,984	16,247	9,814	16,667	44,473	9,877	25,533	44,609
0.1	9,814	10,240	13,154	9,877	3,363	14,688	9,814	20,408	39,698	9,877	26,335	41,259
0.2	9,814	10,240	11,622	9,877	3,363	13,987	9,814	20,408	37,037	9,877	26,335	39,296
0.3	9,814	10,240	10,240	9,877	3,363	13,363	9,814	20,408	34,667	9,877	26,335	38,133
0.4	9,814	10,514	10,514	9,877	1,627	14,127	9,814	23,438	32,544	9,877	26,063	38,379
0.5	9,814	10,514	10,514	9,877	1,627	14,127	9,814	23,438	31,250	9,877	26,063	37,850
0.6	9,814	10,514	10,514	9,877	1,627	14,127	9,814	23,438	29,752	9,877	26,063	37,675
0.7	9,814	10,619	10,619	9,877	1	15,001	9,814	25,926	28,000	9,877	25,141	39,550
0.8	9,814	10,619	10,619	9,877	1	15,001	9,814	25,926	25,926	9,877	25,141	40,141
0.9	9,814	10,619	10,619	9,877	1	15,001	9,814	25,926	25,926	9,877	25,141	40,141
1	9,814	10,619	10,619	9,877	1	15,001	9,814	25,926	25,926	9,877	25,141	40,141

References

Abernathy, William J. and James M. Utterback (1975), "A Dynamic Model of Product and Process Innovation," *Omega*, Volume 3, pages 639-656.

Abernathy, William J. and James M. Utterback (1978), "Patterns of Industrial Innovation," *Technology Review*, Volume 80, pages 41-47.

Abramovitz, Moses and Paul A. David (1973), "Reinterpreting Economic Growth: Parables and Realities," *American Economic Review*, Volume 63, pages 428-439.

Abramovitz, Moses and Paul A. David (1999) "American Macroeconomic Growth in the Era of Knowledge-Based Progress: The Long-Run Perspective," Stanford Institute for Economic Policy Research (SIEPR) Discussion Paper No. 99-3.

Acemoglu, Daron (1996), "A Microfoundations for Social Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics*, Volume 111, pages 779-804.

Acemoglu, Daron (1998), "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics*, Volume 113, pages 1055-1089.

Aghion, Philippe, Nicholas Bloom, Richard Blundell, Rachel Griffith and Peter Howitt (2002), "Competition and Innovation: An Inverted U Relationship," NBER Working Paper No. 9269.

Aghion, Philippe, Christopher Harris, Peter Howitt and John Vickers (2001), "Competition, Imitation and Growth with Step-by-Step Innovation," *Review of Economic Studies*, Volume 68, pages 467-492

Aghion, Philippe and Peter Howitt (1992), "A Model of Growth through Creative Destruction," *Econometrica*, Volume 60, pages 323-351.

Aghion, Philippe and Peter Howitt (1996), "Research and Development in the Growth Process," *Journal of Economic Growth*, Volume 1, pages 49-73.

Aghion, Philippe and Peter Howitt (1998), *Endogenous Growth Theory*, Cambridge, Mass.: MIT Press.

Aghion, Philippe and Peter Howitt (2003), "Growth with Quality-Improving Innovations: An Integrated Framework," draft version in Philippe Aghion and Steven Durlauf (eds.), *Handbook of Economic Growth*, North Holland: Amsterdam, forthcoming.

Ahamdi, Reza H. and Hirofumi Matsuo (1991), "The Line Segmentation Problem," *Operations Research*, Volume 39, pages 42-55.

Amable, Bruno (2002), "Is there an Institutional Base of the New Economy," mimeo.

Amable, Bruno, Rémi Barré and Robert Boyer (1997), *Les Systèmes d'Innovation à l'Ère de la Globalisation*, Paris: Economica.

Amable, Bruno and Pascal Petit (2001), "The Diversity of Social Systems of Innovation and Production During the 1990s," Document de Travail CEPREMAP No. 2001-15.

Antonelli, Christiano (1995), *The Economics of Localized Technological Change and Industrial Dynamics*, Dordrecht and London: Kluwer Academic Publishers.

Arrow, Kenneth J. (1962), "Economic Welfare and the Allocation of Resources for Innovation," in Richard R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton: Princeton University Press.

Arrow, Kenneth J. and William M. Capron (1959), "Dynamic Shortages and Price Rises: The Engineer-Scientist Case," *Quarterly Journal of Economics*, Volume 73, pages 292-308.

Ascher, Uri M. and Rüdiger Weiss (1983), "Collocation for Singular Perturbation Problems I: First Order Systems with Constant Coefficients," *SIAM Journal on Numerical Analysis*, Volume 20, pages 537-557.

Ascher, Uri M. and Rüdiger Weiss (1984), "Collocation for Singular Perturbation Problems II: Linear First Order Systems Without Turning Points," *Mathematics of Computation*, Volume 43, pages 157-187.

Atkinson, Anthony B. and Joseph E. Stiglitz (1969), "A New View of Technological Change," *Economic Journal*, Volume 79, pages 573-578.

Ayres, Robert U. (1968), "Envelope Curve Forecasting," in James R. Bright (ed.), *Technological Forecasting for Industry and Government: Methods and Applications*, New Jersey: Prentice-Hall.

Ayres, Robert U. (1969), *Technological Forecasting and Long-Range Planning*, New York: McGraw-Hill.

Azariadis, Costas and Allan Drazen (1990), "Threshold Externalities in Economic Development," *Quarterly Journal of Economics*, Volume 105, pages 501-526.

Balakrishnan, Anataram and François Vanderbeck (1999), "A Tactical Planning Model for Mixed-Model Electronics Assembly Operations," *Operations Research*, Volume 47, pages 395-409.

Baldwin, Robert E. (1948), "Equilibrium in International Trade: A Diagrammatic Analysis," *Quarterly Journal of Economics*, Volume 62, pages 748-762.

Barro, Robert J. and Xavier Sala-i-Martin (1995), *Economic Growth*, New York: McGraw-Hill.

Baumol, William J. (1982), "Contestable Markets: An Uprising in the Theory of Industry Structure," *American Economic Review*, Volume 72, pages 1-15.

Baumol, William J., John C. Panzar, and Robert D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt Brace Jovanovich.

Bayus, Barry L. (1995), "Optimal Policies for Product and Process Innovation," *Journal of Operations Management*, Volume 12, pages 173-185.

Bayus, Barry L. (1998), "An Analysis of Product Lifetimes in A Technologically Dynamic Industry," *Management Science*, Volume 44, pages 763-775.

Bental, Benjamin and Dan Peled (1996), "The Accumulation of Wealth and the Cyclical Generation of New Technologies: A Search Theoretic Approach," *International Economic Review*, Volume 37, pages 687-718.

Bertola, Giuseppe (1993), "Factor Shares and Savings in Endogenous Growth," *American Economic Review*, Volume 83, pages 1184-1198.

Bishop, Simon and Mike Walker (1999), *The Economics of EC Competition Law: Concepts, Application and Measurement*, London: Sweet and Maxwell.

Bitran, Gabriel R., Elisabeth A. Haas and Hirofumi Matsuo (1986), "Production Planning of Style Goods with High Setup Costs and Forecast Revisions," *Operations Research*, Volume 34, pages 226-236.

Bitros, George C. and Epaminondas E. Panas (1988), "Measuring Product Prices Under Conditions of Quality Change: The Case of Passenger Cars in Greece," *Journal of Industrial Economics*, Volume 37, pages 167-186.

Blitz, Rudolph C. (1958), "Capital Longevity and Economic Development," *American Economic Review*, Volume, pages 313-329.

Brezzi, Monica and Tze Leung Lai (2002), "Optimal Learning and Experimentation in Bandit Problems," *Journal of Economic Dynamics and Control*, Volume 27, pages 87-108.

Bright, James R. (ed.) (1968), *Technological Forecasting for Industry and Government: Methods and Applications*, New Jersey: Prentice-Hall.

Carlsson, Bo and Staffan Jacobsson (1997), "In Search of Useful Public Policies: Key Lessons and Issues for Policy Makers," in Bo Carlsson (ed.), *Technological Systems and Industrial Dynamics*, Boston: Kluwer Academic Publishers.

Caves, Richard E. (1998), "Industrial Organization and New Findings on the Turnover and Mobility of Firms," *Journal of Economic Literature*, Volume 36, pages 1947-1982.

Chamberlin, Edward H. (1948), "Proportionality, Divisibility and Economies of Scale," *Quarterly Journal of Economics*, Volume 62, pages 229-262.

Chiang, Alpha C. (1992), *Elements of Dynamic Optimization*, New York: McGraw Hill.

Cohen, Wesley M. and Steven Klepper (1992), "The Anatomy of Industry R&D Intensity Distributions," *American Economic Review*, Volume 82, pages 773-799.

Cowan, Robin, Paul A. David, and Dominique Foray (2000), "Explicit Economics of Codification and Tacitness," *Industrial and Corporate Change*, Volume 9, pages 211-253.

Cowan, Robin and Dominique Foray (1997), "The Economics of Codification and the Diffusion of Knowledge," *Industrial and Corporate Change*, Volume 6, pages 595-622.

Cowan, Robin and Gert van de Paal (eds.) (1999), "Innovation Policy in a Knowledge-Based Economy," MERIT Memorandum Prepared for the EC DG XIII.

Cowling, Keith and Dennis C. Mueller, "The Social Costs of Monopoly Power," *Economic Journal*, Volume 88, pages 727-748.

Crawford, C. Merle (1992), "The Hidden Costs of Accelerated Product Development," *Journal of Product Innovation Management*, Volume 9, pages 188-199.

Dasgupta, Partha (1988a), "Patents, Priority and Imitation, or, the Economics of Races and Waiting Games," *Economic Journal*, Volume 98, pages 66-80.

Dasgupta, Partha (1988b), "The Welfare Economics of Knowledge Production," *Oxford Review of Economic Policy*, Volume 4, pages 1-12.

Dasgupta, Partha and Paul A. David (1987), "Information Disclosure and the Economics of Science and Technology," in G. Feiwel (ed.), *Arrow and the Ascent of Modern Economic Theory*, New York: New York University Press.

Dasgupta, Partha and Paul A. David (1994), "Towards a New Economics of Science," *Research Policy*, Volume 23, pages 487-521.

Dasgupta, Partha and Joseph E. Stiglitz (1980a), "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, Volume 90, pages 266-293.

Dasgupta, Partha and Joseph E. Stiglitz (1980b), "Uncertainty, Industrial Structure and the Speed of R&D," *Bell Journal of Economics*, Volume 11, pages 1-28.

Dasgupta, Partha and Joseph E. Stiglitz (1988), "Learning-By-Doing, Market Structure and Industrial and Trade Policies," *Oxford Economic Papers*, Volume 40, pages 246-268.

David, Paul A. (1977), "Invention and Accumulation in America's Economic Growth: A Nineteenth Century Parable," in Karl Brunner and Allan H. Meltzer (eds.), *International Organization, National Policies and Economic Development*, Amsterdam: North-Holland.

David, Paul A. (1989), "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror," Center for Economic Policy Research Discussion Paper No. 172, Stanford University.

David, Paul A. (1990a), "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," *American Economic Review*, Volume 80, pages 355-361.

David, Paul A (1990b), "General-Purpose Engines, Investment and Productivity Growth: From the Dynamo Revolution to the Computer Revolution" in Enrico Deiaco, Erik

Hörnell and Graham Vickery (eds.), *Technology and Investment: Crucial Issues for the 1990s*, London: Pinter Publishers.

David, Paul A. (1991), "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror," in OECD, *Technology and Productivity: The Challenge to Economic Policy*, Paris: Organisation for Economic Co-operation and Development.

David, Paul A. (1993a), "Intellectual Property Institutions and the Panda's Thumb: Patents, Copyrights, and Trade Secrets in Economic Theory and History," in Michael B. Wallerstein, Mary Ellen Mogee, and Roberta A. Schoen (eds.), *Global Dimensions of Intellectual Property Rights in Science and Technology*, Washington, DC: National Academy Press.

David, Paul A. (1993b), "Knowledge, Property, and the System Dynamics of Technological Change," in Lawrence H. Summers and Shekhar Shah (eds.), *Proceedings of the World Bank Annual Conference on Development Economics*, Washington, D.C.: World Bank, pages 215-248.

David, Paul A. (1998), "Path Dependence Learning, and the Evolution of Beliefs and Behaviours," forthcoming in Ugo Pagano and Antonio Nicita (eds.), *The Evolution of Economic Diversity*, London: Routledge.

David, Paul A. (2000a), "A Tale of Two Traverses: Continuity and Change in U.S. Macroeconomic History, 1800-1990," [Social Science History Workshop Paper, Department of Economics, Stanford University, 22nd October 1998] revised as Stanford Institute for Economic Policy Research (SIEPR) Discussion Paper.

David, Paul A. (2000b), "Path Dependence and Varieties of Learning in the Evolution of Technological Practice," in John Ziman (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press.

David, Paul A. (2000c), "Understanding the Digital Technology's Evolution and the Path of Measured Productivity Growth: Present and Future in the Mirror of the Past," in Erik Brynjolfsson and Brian Kahin (eds.), *Understanding the Digital Economy*, Cambridge, Mass.: MIT Press.

David, Paul A. (2002), "The Digital Technology Boomerang: New Intellectual Property Rights Threaten Global 'Open Science'," in *Proceedings of the World Bank Annual Conference on Development Economics – Europe 2000*, [CD-ROM Supplement to the World Bank Economic Review], Washington, DC, Summer 2002.

David, Paul A. (2003a), "Koyaanisqatsi in Cyberspace: The Economics of an 'Out-of-Balance' Regime of Private Property Rights in Digital Data and Information," Stanford Institute for Economic Policy Research (SIEPR) Policy Paper No. 02-029.

David, Paul A. (2003b), "Zvi Griliches on Diffusion, Lags and Productivity Growth... Connecting the Dots," Prepared for the Conference on R&D, Education and Productivity Held in Memory of Zvi Griliches (1930-1999), Ministère de la Recherche, Paris, France, 25-27th August 2003, revised as Stanford Institute for Economic Policy Research (SIEPR) Discussion Paper 04-003.

David, Paul A., Bronwyn H. Hall and Andrew A. Toole (2000), "Is Public R&D a Complement or a Substitute for Private R&D? A Review of the Econometric Evidence," *Research Policy*, Volume 29, pages 497-529.

David, Paul A. with J. Gabriel Goddard (2001), "Knowledge, Capabilities and Human Capital Formation in Economic Growth," New Zealand Treasury Working Paper No. 01-13.

David, Paul A. and Trond E. Olsen (1992), "Technology Adoption, Learning Spillovers, and the Optimal Duration of Patent-Based Monopolies," mimeo.

De la Fuente, Angel (2000), *Mathematical Methods and Models for Economists*, Cambridge: Cambridge University Press.

De Matta and Guignard (1994), "Dynamic Production Scheduling for a Process Industry," *Operations Research*, Volume 42, pages 492-503.
Delbono, Flavio (1989), "Market Leadership with a Sequence of History Dependent Patent Races," *Journal of Industrial Economics*, Volume 38, pages 95-101.
Denicolo, Vincenzo (1996), "Patent Races and Optimal Patent Breadth and Length," *Journal of Industrial Economics*, Volume 44, pages 249-265.
Dixit, Avinash K. and Joseph E. Stiglitz (1977), "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, Volume 67, pages 297-308.
Edquist, Charles (ed.) (1997), *Systems of Innovation: Technologies, Institutions and Organizations*, London: Pinter Publishers.
Eicher, Theo S. (1996), "Interaction between Endogenous Human Capital and Technological Change," *Review of Economic Studies*, Volume 63, pages 127-144.
Elliott, John E. (1980), "Marx and Schumpeter on Capitalism's Creative Destruction: A Comparative Restatement," *Quarterly Journal of Economics*, Volume 95, pages 45-68.
Esposito, Emilio and Michele Mastroianni (2002), "Information Technology and Personal Computers: The Relational Life Cycle," *Technovation*, Volume 22, pages 41-50.
Evenson, Robert E. and Yoav Kislev (1975), *Agricultural Research and Productivity*, New Haven: Yale University Press.
Evenson, Robert E. and Yoav Kislev (1976), "A Stochastic Model of Applied Research," *Journal of Political Economy*, Volume 84, pages 265-281.
Filson, Darren (2001), "The Nature and Effects of Technological Change over the Industry Life Cycle," *Review of Economic Dynamics*, Volume 4, pages 460-494.
Filson, Darren (2002), "Product and Process Innovations in the Life Cycle of an Industry," *Journal of Economic Behaviour and Organization*, Volume 1470, pages 1-16.
Fisher, F. M. (1983), *The Disequilibrium Foundations of Equilibrium Economics*, New York: Cambridge University Press.
Francois, Patrick and Huw Lloyd-Ellis (2003), "Animal Spirits Through Creative Destruction," *American Economic Review*, Volume 93, pages 530-550.
Fudenberg, Drew, Richard J. Gilbert, Joseph E. Stiglitz and Jean Tirole (1983) "Preemption, Leapfrogging, and Competition in Patent Races," *European Economic Review*, Volume 22, pages 3-31.
Furske, Mark (2002), *Competition Law of the UK and EC*, 3rd edition, Oxford: Oxford UP.
Gerstenfeld, Arthur (1971), "Technological Forecasting," *The Journal of Business*, Volume 44, pages 10-18.
Gertler, Mark (1988), "Financial Structure and Aggregate Economic Activity: An Overview," *Journal of Money, Credit, and Banking*, Volume 20, pages 559-588.
Gilbert, Richard and Carl Shapiro (1990), "Optimal Patent Length and Breadth," *RAND Journal of Economics*, Volume 21, pages 106-112.
Goolsbee, Austan D (1998), "Does Government R&D Policy Mainly Benefit Scientists and Engineers?," *American Economic Review*, Volume 38, pages 423-428.
Gort, Michael and Steven Klepper (1982), "Time Paths in the Diffusion of Product Innovations," *Economic Journal*, Volume 92, pages 630-653.
Gotz, Georg (2000) "Strategic Timing of Adoption of New Technologies under Uncertainty: A Note," *International Journal of Industrial Organization*, Volume 18, pages 369-379
Griliches, Zvi (1990), "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, Volume 28, pages 1661-1707.
Grossman, G. M. and E. Helpman (1991a), *Innovation and Growth in the Global Economy*, Cambridge, Mass.: MIT Press.

Grossman, G. M. and E. Helpman (1991b), "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, Volume 58, pages 43-61.

Grübler, Arnulf (1990), *The Rise and Fall of Infrastructures: Dynamics of Evolution and Technological Change in Transport*, Heidelberg: Physica-Verlag.

Guderley, Karl G. (1975), "A Unified View of Some Methods for Stiff Two-Point Boundary Value Problems," *SIAM Review*, Volume 17, pages 416-442.

Haberler, Gottfried (1950), "Joseph Alois Schumpeter 1883-1950," *Quarterly Journal of Economics*, Volume 64, pages 333-372.

Hagedoorn, John, Elias Carayannis and Jeffrey Alexander (2001), "Strange Bedfellows in the Personal Computer Industry: Technology Alliances between IBM and Apple," *Research Policy*, Volume 30, pages 837-849.

Hall, Bronwyn H. (2002), "The Assessment: Technology Policy," *Oxford Review of Economic Policy*, Volume 18, pages 1-9.

Hall, Bronwyn H. and John van Reenen (2000), "How Effective are Fiscal Incentives for R&D? A Review of the Evidence Research Policy," *Research Policy*, Volume 29, pages 449-470.

Halman, Johannes I. M., Adrian P. Hoffer and Wim van Vuuren (2003), "Platform-Driven Development of Product Families: Linking Theory with Practice," *Journal of Product Innovation Management*, Volume 20, pages 149-162.

Harberger, Arnold C. (1954), "Monopoly and Resource Allocation," *American Economic Review*, Volume 44, pages 77-87.

Hart, Susan, Erik Jan Hultink, Nikolaos Tzokas and Harry R. Commandeur (2003), "Industrial Companies' Evaluation Criteria in New Product Development Gates," *Journal of Product Innovation Management*, Volume 20, pages 22-36.

Hartl, Richard F., Suresh P. Sethi and Raymond G. Vickson (1995), "A Survey of the Maximum Principles for Optimal Control Problems with State Constraints," *SIAM Review*, Volume 37, pages 181-218.

Hayek, Friedrich A. (1945), "The Use of Knowledge in Society," *American Economic Review*, Volume 35, pages 519-530.

Head, Keith and Thierry Mayer (2001), "Everything you Always Wanted to Know about CES (but Were Afraid to Ask)," mimeo, http://team.univ-paris1.fr/trombi/mayer /ces.pdf.

Helpman, Elhanan and Manuel Trajtenberg (1998a), "A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies," in Elhanan Helpman (ed.), *General Purpose Technologies and Economic Growth*, Cambridge, Mass.: MIT Press.

Helpman, Elhanan and Manuel Trajtenberg (1998b), "Diffusion of General Purpose Technologies," in Elhanan Helpman (ed.), *General Purpose Technologies and Economic Growth*, Cambridge, Mass.: MIT Press.

Hicks, John R. (1932), *The Theory of Wages*, London: Macmillan.

Hicks, John R. (1979), *Causality in Economics*, London: Basil Blackwell.

Hounshell, David A. (1998), "Measuring the Return on Investment in R&D: Voices from the Past, Visions of the Future," in National Research Council, *Assessing the Value of Research in the Chemical Sciences: Report of a Workshop*, Washington: The National Academy Press.

Hounshell, David A. and John K. Smith Jr. (1988), *Science and Corporate Strategy: Du Pont R&D, 1902-1980*, Cambridge: Cambridge University Press.

Jones, Robert and Geoffrey Newman (1995), "Adaptive Capital, Information Depreciation and Schumpeterian Growth," *Economic Journal*, Volume 105, pages 897-915.

Kamien, Morton I. and Nancy L. Schwartz (1972a), "Market Structure, Rivals' Response, and the Firm's Rate of Product Improvement," *Journal of Industrial Economics*, Volume 20, pages 159-172.

Kamien, Morton I. and Nancy L. Schwartz (1972b), "Timing of Innovations Under Rivalry," *Econometrica*, Volume 40, pages 43-60.

Kamien, Morton I. and Nancy L. Schwartz (1975), "Market Structure and Innovation: A Survey," *Journal of Economic Literature*, Volume 13, pages 1-37.

Kamien, Morton I. and Nancy L. Schwartz (1982), *Market Structure and Innovation*, Cambridge University Press: Cambridge.

Keynes, John Maynard (1937), "The General Theory of Employment," *Quarterly Journal of Economics*, Volume 51, pages 209-223.

Klemperer, Paul (1990), "How Broad Should the Scope of Patent Protection Be?," *RAND Journal of Economics*, Volume 21, pages 113-130.

Klepper, Steven (1996), "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *American Economic Review*, Volume 86, pages 562-583.

Klepper, Steven and Elizabeth Graddy (1990), "The Evolution of New Industries and the Determinants of Market Structure," *RAND Journal of Economics*, Volume 21, pages 27-44.

Knight, Frank (1921), *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin.

Knight, Frank (1957), *Risk, Uncertainty and Profit*, 1957 edition, New York: Harper and Row.

Korah, Valentine (1993), "EC Commission Decisions to Competition," in Michael D. A. Freeman and Bob A. Hepple (eds.), *Current Legal Problems 1993*, Volume 46, Part II: Collected Papers, Oxford: Oxford University Press.

Kortum, Samuel S. (1997), "Research, Patenting, and Technological Change," Econometrica, Volume 65, pages 1389-1419.

Kuznets, Simon (1940), "Schumpeter's Business Cycles," *American Economic Review*, Volume 30, pages 257-271.

Kuznets, Simon (1954), *Economic Change: Selected Essays in Business Cycles, National Income, and Economic Growth*, London: Heinemann.

Kuznets, Simon (1966), *Modern Economic Growth: Rate, Structure, and Spread*, New Haven: Yale University Press.

Kuznets, Simon (1971), *Economic Growth of Nations: Total Output and Production Structure*, Cambridge, Mass.: Harvard University Press.

Lakatos, Imre (1970a), "Falsificationism and the Methodology of Scientific Research Programmes," in Imre Lakatos and Alan E. Musgrave, eds, *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.

Lakatos Imre (1970b), "History of Science and Its Rational Reconstructions" (in History of Science and Its Rational Reconstruction), *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1970, pages 91-136.

Léonard, Daniel and Ngo Van Long (1992), *Optimal Control Theory and Static Optimization in Economics*, Cambridge: Cambridge University Press.

LeRoy, Stephen L. and Larry D. Singhell, Jr. (1987), "Knight on Risk and Uncertainty," *Journal of Political Economy*, Volume 95, pages 394-406.

Levine, Ross (1997), "Financial Development and Economic Growth: Views and Agenda," *Journal of Economic Literature*, Volume 35, pages 688-726.

Lipsey, Richard (1998), "Technology Policies in Neo-Classical and Structuralist-Evolutionary Models," *STI Review*, Volume 22, pages 31-73.

Lloyd, Geoffrey E. R. (2002), *The Ambitions of Curiosity*, Cambridge: Cambridge UP, 2002

Loasby, Brian J. (1986), "Organisation, Competition, and the Growth of Knowledge," in Richard N. Langlois (ed.), *Economics As a Process: Essays in the New Institutional Economics*, Cambridge: Cambridge University Press.

Lucas, Robert E., Jr. (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics*, Volume 22, pages 3-42.

Lundvall, Bengt-Åke (ed.) (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London, Pinter Publishers.

Machlup, Fritz (1962), *The Production and Distribution of Knowledge in the United States*, Princeton, Princeton UP.

Machlup, Fritz (1980), *Knowledge: Its Creation, Distribution, and Economic Significance, Volume I, Knowledge and Knowledge Production*, Princeton: Princeton UP.

Mahajan, Vijay and Eitan Muller (1996), "Timing, Diffusion, and Substitution of Successive Generations of Technological Innovations: The IBM Mainframe Case," *Technological Forecasting and Social Change*, Volume 51, pages 109-132.

Malerba, Franco (1998), "Public Policy and Industrial Dynamics: An Evolutionary Perspective," ISE Report Sub-Project 3.1.1, ISE CD-ROM, Systems of Innovation Research Programme, Department of Technology and Social Change, Linköping University.

Malerba, Franco, Richard Nelson, Luigi Orsenigo and Sydney Winter (1999), "'History-Friendly' Models of Industry Evolution: The Computer Industry," *Industrial and Corporate Change*, Volume 8, pages 3-40.

Malerba, Franco, Richard Nelson, Luigi Orsenigo and Sydney Winter (2001a), "History-Friendly Models: An Overview of the Case of the Computer Industry," *Journal of Artificial Societies and Social Simulation*, Volume 4, http://www.soc.surrey.ac.uk/JASSS/4/3/6.html.

Malerba, Franco, Richard Nelson, Luigi Orsenigo and Sydney Winter (2001b), "Competition and Industrial Policies in a History-Friendly Model of the Evolution of the Computer Industry," *International Journal of Industrial Organization*, Volume 19, pages 635-644.

Mansfield, Edwin (1985), "How Rapidly Does New Industrial Technology Leak Out?," *Journal of Industrial Economics*, Volume 34, pages 217-223.

Mansfield, Edwin (1991), "Academic Research and Industrial Innovation," *Research Policy*, Volume 20, pages 1-12

Mansfield, Edwin (1995), "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing," *Review of Economics and Statistics*, Volume 77, pages 55-65.

Mansfield, Edwin (1998), "Academic Research and Industrial Innovation: An Update of Empirical Findings," *Research Policy*, Volume 26, pages 773-776.

Mansfield, Edwin, Mark Schwartz and Samuel Wagner (1981), "Imitation Costs and Patents: An Empirical Study," *The Economic Journal*, Volume 91, pages 907-918.

Marey, Philip and Lex Borghans (2000), "Wage Elasticities of the Supply of Knowledge Workers in the Netherlands," Research Centre for Education and the Labour Market Paper ROA-R-2000/6E, Maastricht University.

Marshall, Alfred (1956), *Principles of Economics: An Introductory Volume*, 8th edition, London: MacMillan and Company.

Martin, Stephen (2000), "Competition Policy for High Technology Industries," Tinbergen Institute Discussion Paper No. 00-010/2.

Martino, Joseph P. (1986), *Technological Forecasting and Decision Making*, New York: North-Holland.

Matsumoto, Kiyofumi, Noritomo Ouchi, Chihiro Watanabe and Charla Griffy-Brown (2002), "Optimal Timing of the Development of Innovative Goods with Generation – An Empirical Analysis Focusing on Canon's Printer Series," *Technovation*, Volume 22, pages 175-185.

Metcalfe, J. Stanley (1994), "Evolutionary Economics and Technology Policy," *Economic Journal*, Volume 104, pages 931-944.

Metcalfe, J. Stanley (1995), "Technology Systems and Technology Policy in an Evolutionary Framework," *Cambridge Journal of Economics*, Volume 19, pages 25-46.

Metcalfe, J. Stanley and Luke Georghiou (1998), "Equilibrium and Evolutionary Foundations of Technology Policy," *STI Review*, Volume 22, pages 75-100.

Meyer, Marc H. and Dhaval Dalal (2002), "Managing Platform Architectures and Manufacturing Processes for Nonassembled Products," *Journal of Product Innovation Management*, Volume 19, pages 277–293.

Meyer, Marc H. and Luis Lopez (1995), "Technology Strategy in a Software Products Company," *Journal of Product Innovation Management*, Volume 12, pages 294-306.

Meyer, Marc H., Peter Tertzakian and James M. Utterback (1997), "Metrics for Managing Research and Development in the Context of the Product Family," *Management Science*, Volume 43, pages 88-111.

Minehart, Deborah and Suzanne Scotchmer (1999), "Ex Post Regret and the Decentralized Sharing of Information," *Games and Economic Behavior*, Volume 27, pages 114-131.

Muffato, Moreno and Marco Roveda (2000), "Developing Product Platforms: Analysis of the Development Process," *Technovation*, Volume 20, pages 617-630.

Mulkay, Michael J. (1972), *The Social Process of Innovation: A Study in the Sociology of Science*, London: Macmillan Press.

Narin, Francis, Kimberly S. Hamilton and Dominic Olivastro (1997), "The Increasing Linkage between U.S. Technology and Public Science," *Research Policy*, Volume 26, pages 317-330.

National Research Council (1999), *Funding a Revolution: Government Support for Computing Research*, Washington, D.C.: National Academy Press.

Nelson, Richard R. (ed.) (1993), *National Innovation Systems: A Comparative Analysis*, Oxford: Oxford University Press.

Nelson, Richard R. and Edmund S. Phelps (1966), "Investment in Humans, Technological Diffusion, and Economic Growth," *American Economic Review*, Volume 56, pages 69-75.

Nelson, Richard R. and Sydney G. Winter (1974), "Neoclassical vs. Evolutionary Theories of Economic Growth: Critique and Prospectus," *Economic Journal*, Volume 84, pages 886-905.

Nelson, Richard R. And Sydney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, Mass.: Belknap Press of Harvard University Press.

Nelson, Richard R. and Sidney G. Winter (2002), "Evolutionary Theorizing in Economics," *Journal of Economic Perspectives*, Volume 16, pages 23-46.

Nordhaus, William D. (1967), "The Optimal Life of a Patent," Cowles Foundation Discussion Paper No. 241.

Nordhaus, William D. (1969), *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change*, Cambridge, Mass.: M.I.T. Press.

Nordhaus, William D. (1972), "The Optimum Life of a Patent: Reply" (in Communications), *American Economic Review*, Volume 62, pages 428-431.

Northrop, Linda M. (2002), "SEI's Software Product Line Tenets," *IEEE Software*, Volume 19, pages 32-40.

OECD (1994), *Accessing and Expanding the Science and Technology Knowledge Base*, Paris: Organisation for Economic Co-operation and Development.

OECD (1995), *National Systems for Financing Innovation*, Paris: Organisation for Economic Co-operation and Development.

OECD (1998), *Technology, Productivity and Job Creation: Best Policy Practice*, Paris: Organisation for Economic Co-operation and Development.

Ordover, Janusz A. and William Baumol (1988), "Antitrust Policy and High-Technology Industries," *Oxford Review of Economic Policy*, Volume 4, pages 13-34

Osawa, Yoshitaka (2003), "How Well Did the New Sumitomo Electric Project Ranking Method Predict Performance," *R&D Management*, Volume 33, pages 343-350.

Pagano, Marco (1993), "Financial Markets and Growth: An Overview," *European Economic Review*, Volume 37, pages 613-622.

Perelman, Michael (1995), "Retrospectives: Schumpeter, David Wells, and Creative Destruction," *Journal of Economic Perspectives*, Volume 9, pages 189-197.

Perez, Carlota and Luc Soete (1988), "Catching Up in Technology: Entry Barriers and Windows of Opportunity," in Giovanni Dosi et al. (eds.), *Technical Change and Economic Theory*, London: Pinter.

Pleatsikas, Christopher and David Teece (2001), "The Analysis of Market Definition and Market Power in the Context of Rapid Innovation," *International Journal of Industrial Organization*, Volume 19, pages 665-693.

Pugh, Emerson W. (1995), *Building IBM: Shaping an Industry and its Technology*, Cambridge, Mass.: MIT Press.

Ramsey, Frank (1928), "A Mathematical Theory of Saving," *Economic Journal*, Volume 38, pages 543-559.

Redding, S. (1996), "The Low-Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D," *Economic Journal*, Volume 106, pages 458-470.

Rigotti, Luca (2003), "Decisive Entrepreneurs and Cautious Investors," mimeo, http://faculty.fuqua.duke.edu/~rigotti/bio/finance.pdf.

Roberts, Rhonda (1998), "Managing Innovation: The Pursuit of Competitive Advantage and the Design of Innovation Intense Environments," *Research Policy*, Volume 27, pages 159-175.

Robinson, Joan (1962), *Essays in the Theory of Economic Growth*, London: MacMillan and Company.

Romer, Paul M. (1986), "Increasing Returns and Long-run Growth," *Journal of Political Economy*, Volume 94, pages 1002-1037.

Romer, Paul M. (1990), "Endogenous Technological Change," *Journal of Political Economy*, Volume 98, pages S71-102.

Rosenberg, Nathan (1974), "Science, Invention and Economic Growth," *Economic Journal*, Volume 84, pages 90-108.

Rosenberg, Nathan (1996), "Uncertainty and Technological Change," in Ralph Landau, Timothy Taylor and Gavin Wright (eds.), *The Mosaic of Economic Growth*, Stanford, Ca.: Stanford University Press.

Rostow, Walt W. (1975), "Kondratieff, Schumpeter, and Kuznets: Trend Periods Revisited," *Journal of Economic History*, Volume 35, pages 719-753.

Runde, Jochen (1998), "Clarifying Frank Knight's Discussion of the Meaning of Risk and Uncertainty," *Cambridge Journal of Economics*, Volume 22, pages 539-546.

Russell, R. D. and J. Christiansen (1978), "Adaptive Mesh Selection Strategies for Solving Boundary Value Problems," *SIAM Journal on Numerical Analysis*, Volume 15, pages 59-80.

Sahal, Devendra (1985), "Technological Guideposts and Innovation Avenues," *Research Policy*, Volume 14, pages 61-82.

Sanderson, Susan and Mustafa Uzumeri (1995), "Managing Product Families: The Case of the Sony Walkman," *Research Policy*, Volume 24, pages 761-782.

Scherer, F. M. (1992), "Schumpeter and Plausible Capitalism," *Journal of economic Literature*, Volume 30, pages 1416-1433.

Schmookler, Jacob (1954), "The Level of Inventive Activity," *Review of Economics and Statistic*, Volume 36, pages 183-190.

Schmookler, Jacob (1962), "Economic Sources of Inventive Activity," *Journal of Economic History*, Volume 22, pages 1-20.

Schmookler, Jacob (1966), *Invention and Economic Growth*, Cambridge, Mass.: Harvard University Press.

Schumpeter, Joseph A. (1942), *Capitalism, Socialism and Democracy*, New York: Harper and Brothers.

Schumpeter, Joseph A. (1947), *Capitalism, Socialism, and Democracy*, 2nd edition, New York: Harper and Brothers.

Schumpeter, Joseph A. (1934), *The Theory of Economic Development*, Cambridge, Mass.: Harvard University Press.

Schumpeter, Joseph A. (1939), *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, New York: McGraw-Hill.

Scott, Maurice F. (1989), *A New View of Economic Growth*, Oxford: Clarendon Press.

Seierstad, Atle and Knut Sydsaeter (1977), "Sufficient Conditions in Optimal Control Theory," *International Economic Review*, Volume 18, pages 367-391.

Shafer, Glenn (1976), *A Mathematical Theory of Evidence*, Princeton: Princeton University Press.

Shafer, Glenn (1990), "Perspectives on the Theory and Practice of Belief Functions, *International Journal of Approximate Reasoning,* Volume 3, pages 1-40.

Shampine, Lawrence F., Jacek Reichelt, and Mark W. Kierzenka (2000), "Solving Boundary Value Problems for Ordinary Differential Equations in MATLAB with bvp4c," ftp://ftp.mathworks.com/pub/doc/papers/bvp/

Simmel, Georg (1990), *The Philosophy of Money*, 2nd edition, Translated by Thomas Bottomore and David Frisby, London: Routledge.

Soete, Luc and Anthony Arundel (eds.) (1993), "An Integrated Approach to European Innovation and Technology Diffusion Policy: A Maastricht Memorandum," Commission of the European Communities Publication No. EUR 15090 EN.

Solow, Robert M. (1956), "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, Volume 70, pages 65-94.

Solow, Robert M. (1957), "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, Volume 39, pages 312-320.

Solow, Robert M. (1994), "Perspectives on Growth Theory," *Journal of Economic Perspectives*, Volume 8, pages 45-54.

Stein, Jeremy C. (1997), "Waves of Creative Destruction: Firm-Specific Learning-by-Doing and the Dynamics of Innovation," *Review of Economic Studies*, Volume 64, pages 265-288.

Stephan, Paula E. (1996), "The Economics of Science," *Journal of Economic Literature*, Volume 34, pages 1199-1235.

Stiglitz, Joseph E. (1987), "Technological Change, Sunk Costs, and Competition," *Brookings Papers on Economic Activity*, Volume 3, pages 883-947.

Stokey, Nancy L. (1991), "Human Capital, Product Quality, and Growth," *Quarterly Journal of Economics*, Volume 106, pages 587-616.

Stokey, Nancy L. (1995), "R&D and Economic Growth," *Review of Economic Studies*, Volume 62, pages 469-489.
Stolper, Wolfgang F. (1951), "The Schumpeterian System" (in Review Articles and Discussions), *Journal of Economic History*, Volume 11, pages 272-277.
Stoneman, Paul and Paul Diederen (1994), "Technology Diffusion and Public Policy," *Economic Journal*, Volume 104, pages 918-930.
Strassmann, W. Paul (1959), "Creative Destruction and Partial Obsolescence in American Economic Development," *Journal of Economic History*, Volume 19, pages 335-349.
Sutton, John (1991), *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*, Cambridge, Mass.: MIT Press.
Sutton, John (1997), "One Smart Agent," *RAND Journal of Economics*, Volume 28, pages 605-628.
Sutton, John (1998), *Technology and Market Structure: Theory and History*, Cambridge, Mass.: MIT Press.
Swan, T. W. (1956), "Economic Growth and Capital Accumulation," *Economic Record*, Volume 32, pages 334-361.
Sweezy, Paul M. (1943), "Professor Schumpeter's Theory of Innovation," *Review of Economic Statistics*, Volume 25, pages 93-96.
Symeonidis, George (1999), "Cartel Stability in Advertising-Intensive and R&D-Intensive Industries," *Economics Letters*, Volume 62, pages 121-129.
Symeonidis, George (2000), "Price and Non-Price Competition with Endogenous Market Structure," *Journal of Economics and Management Strategy*, Volume 9, pages 53-83.
Symeonidis, George (2003), "Quality Heterogeneity and Welfare," *Economics Letters*, Volume 78, pages 1-7.
Tatikonda, Mohan V. (1999), "An Empirical Study of Platform and Derivative Product Development Projects," *Journal of Product Innovation Management*, Volume 16, pages 3-26.
Teece, David and Mary Coleman (1998), "The Meaning of Monopoly: Antitrust Analysis in High Technology Industries," *Antitrust Bulletin*, Volume 43, pages 801-857.
Telser, Lester G. (1982), "A Theory of Innovation and Its Effects," *RAND Journal of Economics*, Volume 13, pages 69-92.
Thoreau, Henry D. (1999), *Walden*, Oxford: Oxford University Press.
Tidd, Joe and Kirsten Bodley (2002), "The Influence of Project Novelty on the New Product Development Process," *R&D Management*, Volume 32, pages 127-138.
Trajtenberg, M., R. Henderson, and A. Jaffe (1997), "University versus corporate patents: A window on the basicness of invention," *Economics of Innovation and New Technology*, Volume 5, pages 19-50.
Tyndall, John (1860), *The Glaciers of the Alps: A Narrative of Excursions and Ascents*, London.
Uzumeri, Mustafa and Susan Sanderson (1995), "A Framework for Model and Product Family Competition," *Research Policy*, Volume 24, pages 583-607.
Vaithianathan, Rhema, Luca Rigotti and Matthew Ryan (2001), "Entrepreneurial Innovation," UC Berkeley Department of Economics Working Paper No. E01-296.
Vernon, Raymond (1966), "International Investment and International Trade in the Product Cycle," *Quarterly Journal of Economics*, Volume 80, pages 190-207.
Verspagen, Bart (2002), "Structural Change and Technology. A Long View," Eindhoven Centre for Innovation Studies Working Paper No. 02.13, Eindhoven University of Technology.
Vickers, John (1986), "The Evolution of Market Structure When There Is a Sequence of Innovations," *Journal of Industrial Economics*, Volume 35, pages 1-12.

Vickers, John (1995), "Concepts of Competition," *Oxford Economic Papers*, Volume 47, pages 1-23.

Vind, Karl (1967), "Control Systems with Jumps in the State Variables," *Econometrica*, Volume 35, pages 273-277.

Whish, Richard (1993), *Competition Law*, 3rd edition, London: Butterworths.

Wilson, Lynn O. and John A. Norton (1989), "Optimal Entry Time for a Product Line Extension," *Marketing Science*, Volume 8, pages 1-17.

Witherspoon, Gary (1977), *Language and Art in the Navajo Universe*, Ann Arbor: University of Michigan Press.

Worrall, John (1998), "Lakatos, Imre," in Edward Craig (ed.), *Routledge Encyclopedia of Philosophy*, London: Routledge. Retrieved September 24, 2003, from http://www.rep.routledge.com/article/DD035SECT3

Young, Alwyn (1991), "Learning by Doing and the Dynamic Effects of International Trade," *Quarterly Journal of Economics*, Volume 106, pages 369-405.

Young, Alwyn (1993), "Substitution and Complementarity in Endogenous Innovation," *Quarterly Journal of Economics*, Volume 108, pages 775-807.

Zeckhauser, Richard (1968), "Optimality in a World of Progress and Learning," *Review of Economic Studies*, Volume 35, pages 363-365.