Knowledge, growth without scale effects, and the product life cycle

André van Hoorn
Tilburg
December 2003

Knowledge, growth without scale effects, and the product life cycle

Master's Thesis
Department of Economics
Tilburg University

Supervisor: Dr. S. Smulders

André van Hoorn
Tilburg
December 2003

Preface/Acknowledgements

I would like to use this space to acknowledge the contributions of the people that have significantly to the quality of this thesis and its realization.

First of all, I would like to thank Dr. S. Smulders who was willing to supervise my thesis efforts. My interest in economics in general and in the subject of (endogenous/R&D-based) economic growth in particular, truly sparked when I attended his course on 'Growth and Technology'. This course left me with several questions, especially regarding the link between the observed reality of economic growth and the predictions of R&D-based models of growth. In this thesis, I have actually sought to answer some of those questions. For myself, I have at times succeeded to do so, for which Dr. S. Smulders deserves a lot of credit. Discussing my results with him always provided me with useful comments and new perspectives on an issue. Especially there where my line of reasoning lacked much-needed nuance, he provided me with useful additional information.

Second, I would like to thank Prof. van Schaik. His seminar on 'Trade and Growth' (Seminar International Economics) furthered my interest in the subject of economic growth and he was also willing to take place on my examination committee.

Finally, I would like to thank all the people who have somehow, someway, contributed to this thesis though in a less visible way.

Contents

CHAPTER 1 INTRODUCTION 2

CHAPTER 2 GROWTH AND SCALE EFFECTS 4
2.1 KNOWLEDGE AND SCALE EFFECTS IN GROWTH 4
2.2 EMPIRICAL EVIDENCE 6
2.2.1 Suitable measures of size 6
2.2.2 Facts on GDP and population 8
2.2.3 The Jones evidence 9
2.2.4 Other evidence 11
2.2.5 Concluding remarks 15
2.3 MODELS OF GROWTH WITHOUT SCALE EFFECTS 16
2.3.1 One-R&D-sector semi-endogenous growth models 16
2.3.2 Two-R&D-sector endogenous growth models 20
2.3.3 Recent models of growth without scale effects 23
2.3.4 Concluding remarks 26

CHAPTER 3 KNOWLEDGE, R&D AND SPILLOVERS, AT THE FIRM LEVEL 29
3.1 FIRM SIZE AND GROWTH 30
3.2 R&D AND PURE KNOWLEDGE SPILLOVERS 36
3.2.1 Country-level knowledge production 38
3.2.2 Interfirm pure knowledge spillovers 45
3.2.3 Intrafirm pure knowledge spillovers 47
3.2.4 Concluding remarks 48
3.3 KNOWLEDGE, R&D, SPILLOVERS, INCREASING RETURNS, AND THE FIRM 52
3.3.1 R&D and diversification 53
3.3.2 R&D and firm size 57
3.3.3 Concluding remarks 60
3.4 DIVERSIFICATION AND FIRM PERFORMANCE 63

3.4.1 Motives for diversification 64
3.4.2 Empirical work 68
3.4.3 Theoretical work 74
3.4.4 Concluding remarks 77
3.5 FIRM SIZE AND FIRM PERFORMANCE 81
3.6 DISCUSSION AND CONCLUSIONS 83

CHAPTER 4 GROWTH WITHOUT SCALE EFFECTS AND STRUCTURAL CHANGE 90
4.1 A SIMPLE MODEL 90
4.1.1 The basic framework 91
4.1.2 A three-sector-model 97
4.1.3 The three-sector-model with spillovers 100
4.1.4 Concluding remarks 103
4.2 GROWTH, INCOME, AND DEMAND 105
4.3 STRUCTURAL CHANGE AND DEINDUSTRIALIZATION 111
4.3.1 Data on structural change 111
4.3.2 Growth, trade, and deindustrialization 113
4.3.3 Structural change and growth 117
4.4 DISCUSSION AND CONCLUSIONS 120

CHAPTER 5 MEASUREMENT ISSUES IN THE STUDY OF R&D-BASED GROWTH 125
5.1 TFP AND THE KNOWLEDGE STOCK 126
5.1.1 Definition and use of knowledge in R&D-based models of growth 127
5.1.2 The measurement of knowledge 132
5.2 OUTPUT, GROWTH, THE CPI, AND WELL-BEING 135
5.2.1 The CPI, output, and growth 135
5.2.2 Income, well-being, and growth 143
5.3 DISCUSSION AND CONCLUSIONS 147
5.3.1 Knowledge, knowledge production, obsolescence, and creative destruction 148

5.3.2 The measurement of output and growth scale effects 151
5.3.3 Well-being over time and creative destruction 153
5.3.4 Concluding remarks 155

CHAPTER 6 THE PRODUCT LIFE CYCLE, DEMAND, TECHNOLOGICAL PROGRESS, AND GROWTH WITHOUT SCALE EFFECTS 157
6.1 THE LIFE CYCLE 157
6.1.1 The diffusion of technology 158
6.1.2 The geography of production and trade 162
6.1.3 The evolution of industries 163
6.2 DEMAND AND GROWTH WITHOUT SCALE EFFECTS 166
6.2.1 Hierarchic demand, saturation, and nonhomothetic preferences 166
6.2.2 Theoretical work with hierarchic demand 168
6.3 DISCUSSION AND CONCLUSIONS 171

CHAPTER 7 CONCLUDING REMARKS 173

APPENDIX A: Rent spillovers and pure knowledge spillovers 179
APPENDIX B: Firm size and growth 182
APPENDIX C: Pure knowledge spillovers at the firm level 185
APPENDIX D: The diversification-performance nexus 197
APPENDIX E: Sectoral structure, value added per worker, and growth 206
APPENDIX F: Growth, schooling, and income 209
APPENDIX G: Sectoral structure and growth 211
APPENDIX H: Growth scale effects in measurable sectors 214
APPENDIX I: S-shaped vs. hump-shaped 216

REFERENCES

LIST OF TABLES

TABLE 1:	Main features of some work on the growth-size relation	33
TABLE 2:	Country-level knowledge production	39
TABLE 3:	Main features of some recent work on the diversification-performance nexus	71
TABLE 4:	Output shares and sectoral growth rates	113
TABLE 5:	Explaining deindustrialization	115
TABLE 6:	Structural change and growth	117
TABLE 7:	Quality-change new-product bias for selected CPI (sub)components over time	139
TABLE 8:	Industry evolution	165

LIST OF ILLUSTRATIONS

FIGURE 1:	GDP per Capita 1870-1987	8
FIGURE 2:	Population 1870-1987	9
FIGURE 3:	Scientists and engineers engaged in R&D	10
FIGURE 4:	Profitability of independent firms i and j and merged firms i and j	78
FIGURE 5:	The growth rate of sectoral TFP over time	92
FIGURE 6:	A numerical example of a simple two-sector-model of growth with sectoral demand shifts	95
FIGURE 7:	Growth rate of total output in the simple two-sector-model	96
FIGURE 8:	A numerical example of a simple three-sector-model of growth with sectoral demand shifts	98
FIGURE 9:	Growth rate of total output in the simple three-sector-model	99
FIGURE 10:	A numerical example of a simple three-sector-model of growth with sectoral demand shifts and spillovers	101
FIGURE 11:	Demand for products 1 and 2 over time in the three-sector-model	104
FIGURE 12:	The typical S-shaped life cycle with a logistic demand curve	109
FIGURE 13:	Sectoral structure of total output/demand 1950-1990	112
FIGURE 14:	Growth, schooling, and income	118
FIGURE 15:	Value added per worker for the U.S. measurable sectors and a	

hypothetical economy 141
FIGURE 16: Sector size and sectoral growth rate for the Agricultural sector 142
FIGURE 17: Percentage of use (market penetration) as a function of time 159
FIGURE 18: Diffusion of new - and replacement of old - products 161
FIGURE 19: Market penetration over the life cycle 161
FIGURE 20: The geographical life cycle for U.S. production and trade 162
FIGURE 21: Entry, exit, and the total number of producers over the life cycle 164
FIGURE 22: The Flying Geese pattern of different product life cycles 169

Abstract

In the early models of R&D-based growth, knowledge's nonrivalrous nature gives rise to spillovers which in turn lead to a scale effect, i.e. larger economies, suitably measured, grow faster than smaller economies do. Empirical work has, however, pointed out that growth with scale effects as predicted by these early models is inconsistent with observed economic reality in which over the last 40 years the OECD countries have experienced a tremendous rise in population, total GDP and the number of people involved in R&D activities whereas the growth rates of per-capita income have shown no corresponding increase and have actually remained relatively constant. This observation led to new models of R&D-based growth in which growth scale effects are removed by limiting the extent of the spillovers associated with knowledge's nonrivalrousness. Generally, however, this is done in an unearned way that lacks much-needed (micro-)economic foundation. This thesis' objective is to gather clues that help further the understanding of (the mechanism underlying) the observed absence of growth scale effects and enhance the ability to credibly model growth without scale effects. For this purpose it takes a closer look at empirical work that reports on regularities and phenomena at both the micro- and the macro level which are expected to provide insight on this mechanism.

The tenor of the empirical work is that spillovers are existent and that knowledge is (at least partly) nonrivalrous (but growth scale effects are absent nonetheless). However, there is also evidence that (i) spillovers resulting from knowledge's nonrivalrous nature are diluted through technological distance, and that (ii) life-cycle effects are important, i.e. (many) products are subjected to a life cycle with the well-known phases of take-off, maturity and, eventually, decline.

Life-cycle effects form the main mechanism behind the proposed solution to reconciling the seemingly contradictory findings of knowledge being (at least partly) nonrivalrous and giving rise to spillovers with the fact that there is no scale effect in growth. The idea is that over the life cycle of a product growth might actually exhibit a scale effect, i.e. the rate at which the ability of producing the product grows, increases over time. That growth at the aggregate level nonetheless does not increase over time is due to two important effects: (1) over time demand (consumption) continuously shifts from one product to another, and (2) the economy's productivity in newly consumed products is lower than in the previously consumed products due to technological distance, which dilutes the knowledge stock firms have built with regard to the production of the previously consumed products, i.e. knowledge is at least partly product-specific. That knowledge used in the production of previously consumed goods has only limited usefulness in the production of currently consumed goods means that (product-specific) knowledge loses its value (i.e. depreciates or becomes obsolete) as demand (and price) for the associated product decline(s). Ultimately the result is that growth at the aggregate level (as measured by output increases measured in monetary units) follows a steady path with a growth rate that is roughly constant over time.

Important in this mechanism, besides the continuous shift of demand to new (recently introduced) goods, is the complex interaction between factors like consumer preferences, demand, income, technological progress and well-being. People will only shift demand towards new (more expensive) products if this increases their well-being. In this way, growth may not exhibit a scale effect while well-being does. However, the relation between income, demand and well-being is such that new products do not only influence the level of well-being but also somehow diminish well-being, for instance by diminishing the utility people derive from the previously consumed (older) products. Consumption of the newly introduced products might then be needed just to keep the level of well-being constant. Work that has shown that for several countries the level of well-being has not increased, despite enormous increases of per-capita income, confirms this idea.

The reason why growth does not exhibit a scale effect at the aggregate level consequently lies in what is here labeled 'the endogeneity of the product life cycle': the factors that give rise to the life cycle, e.g. productivity growth, monotonically declining prices, shifting preferences, the level of income etc., can all be traced back to a two-way causality between demand and technological progress in a broad sense. Demand factors influence firms' R&D efforts and consequently technological progress, i.e. productivity growth and the introduction of new and/or improved products, in turn affecting demand and indeed future R&D activities, and along the way giving rise to several forms of creative destruction, e.g. the obsolescence of knowledge specific to products no longer in demand. By distancing themselves from the standard Dixit-Stiglitz production function (Dixit and Stiglitz 1977) (i.e. homothetic preferences) and instead using so-called hierarchic (nonhomothetic) preferences several researchers have not only endogenized the product life cycle, but also have provided a framework that could be used to model growth without scale effects more credibly than the currently existing models do.

1 Introduction

For several reasons, knowledge cannot be treated like any other commodity. One of these reasons is the nonrivalrous nature of knowledge, which means that one person's use of certain knowledge does not diminish another person's use of the same knowledge (at the same time). This important property of knowledge is used in several early models of R&D-based growth[1], e.g. Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992). In these models this property leads to a scale effect, which boils down to larger economies growing faster than smaller economies (with the measure of size suitably defined (cf. Backus, Kehoe and Kehoe 1992)).

In an influential paper, Jones (1995a) pointed out that growth with scale effects, as predicted by the early models of R&D-based growth, is inconsistent with empirical facts. Over the last 40 years the OECD countries have experienced a tremendous rise in the number of people involved in R&D activities whereas the growth rates of per-capita income have shown no corresponding increase. This is a puzzling observation and has led to new models of R&D-based growth that did not incorporate scale effects e.g. Jones (1995b), Smulders and van de Klundert (1995), Young (1998), Li (2000), and Peretto and Smulders (2002).

Generally, however, these models suffer from the 'Solow critique'; Solow (1994) criticizes (some) growth theorists because they "… often just insert favorable assumptions in an unearned way; and then when they put in their thumb and pull out the vary plum they have inserted, there is a tendency to think that something has been proved." (p. 53). In the models of growth without scale effects the prediction of a scale effects in growth of the early models of R&D-based growth is removed by limiting the extent of the spillovers associated with knowledge's nonrivalrousness, but often the much-needed (micro-)economic foundation for the crucial assumption in these models regarding the extent of knowledge spillovers - and the mechanism limiting their extent - is lacking. Assuming that knowledge is rivalrous (not nonrivalrous) to limit spillovers and dispose of the scale effects prediction of the early models of R&D-based growth simply does not shed much light on the issue of growth without scale effects however.

[1] R&D-based growth models are also often referred to as endogenous growth models or Schumpeterian growth models. The most important feature of these models, be they R&D-based, endogenous or Schumpeterian, is that knowledge or the knowledge stock does not increase exogenously as in the neoclassical model of growth (e.g. Solow 1956), but endogenously as the result of (profit) maximizing behavior of economic subjects.

This thesis can subsequently be seen as an attempt to address the Solow critique for models of (R&D-based) growth without scale effects. Its objective is simply to gather clues that help further the understanding of (the mechanism underlying) the observed absence of growth scale effects and (consequently) enhance the ability to credibly model growth without scale effects. For this, a closer look is taken at available documentation on actually observed regularities and phenomena at both the micro and the macro level. The clues that are picked up along the way point to important aspects of economic reality that could and should play an important role in explaining the absence of a scale effect in growth where logically such an effect would be expected as the result of knowledge's nonrivalrousness.

The thesis continues with a discussion of growth and scale effects in Chapter 2. In this chapter explicit attention will be paid to the empirical evidence on the existence (and absence) of scale effects in growth and to some theoretical models of growth without scale effects. Chapter 3 examines work that has investigated the role factors that play an important role in the study of aggregate-level growth and scale effects, namely knowledge, R&D and spillovers, at the firm level. The idea is that gaining more insight in the role these factors play at the micro level, will provide information useful for the understanding of - and the ability to credibly model - growth without scale effects. In Chapter 4, a simple dynamic explanation for the absence of scale effects is presented. The idea is that the contribution to GDP and output (productivity) growth of sectors (or industries or products), which actually do exhibit a scale effect in growth, gets lost in the aggregate. This idea is substantiated by models incorporating structural change and empirical evidence on deindustrialization, a manifestation of structural change. Chapter 5 subsequently discusses some measurement issues that are bound to surface in the study of R&D-based growth. Besides the fact that some understanding of these issues is useful in itself, the discussion in this chapter also sheds light on the mechanism that causes the absence of growth scale effects (where logically such effects are expected as the result of knowledge's nonrivalrous nature). Chapter 6 elaborates on the results of the previous two chapters and ultimately concludes that some sort of endogenous product life cycle plays an important role in growth without scale effects. In Chapter 7 some concluding remarks including some suggestions for further research are presented, and finally, the appendices provide background information regarding, amongst others, work discussed in the main text, data used in figures, etc.

2 Growth and scale effects

2.1 Knowledge and scale effects in growth

Before discussing the empirics and theoretical models of growth without scale effects, it seems sensible to elaborate further on knowledge and its unique, compared to more tangible productive resources like (physical) capital, properties that form the foundation for scale effects in growth.
For now knowledge is defined as information available to a person (or persons) that determines his (or her)[2] (their) ability to combine resources to produce goods. In addition, knowledge and technology are used as synonyms.

Pertaining to scale effects in growth, the most important property of knowledge is its nonrivalrous nature (see for instance Romer 1990, and Nelson 1959, Arrow 1962 and Nordhaus 1969 for earlier discussions). The fact that knowledge is nonrivalrous means that one person's use of certain knowledge does not diminish another person's use of the same knowledge (at the same time). Ultimately, this nonrivalrousness leads to increasing returns to scale. The costs of creating new knowledge are relatively high. These cost are, however, independent of its subsequent use. Once new knowledge for example in the form of a new product has been created, this knowledge can be used over and over again in the production of more units of this new product without having to create the required knowledge again. The costs of subsequent units are consequently relatively small vis-à-vis the costs of creating the knowledge and the product for the first time. Hence, the production function for products that use newly created knowledge is such that it exhibits increasing returns. This notion can best be clarified by an example.[3] Hungarian born Erno Rubik, a lecturer at the Department of Interior Design at the Academy of Applied Arts and Crafts in Budapest, spend several years developing his maddening puzzle, the Cube, when in 1974 he created the first working

[2] Where in this thesis "his" or "he" or their related pronouns appear they are used in their generic sense to include both male and female sexes.
[3] In addition, a more formal description of the increasing returns associated with the nonrivalrous nature of knowledge is that, knowledge related to product i can be used in the production of any given number of product i and/or the production of any given number of products of type j (or k, l, etc.) at relatively small costs compared to the costs associated with its initial creation.

prototype and drafted a patent application.[4] This was, however, only the first step in the process that would ultimately lead to the staggering sales numbers and massive success of the cube. In fact, there would not have been a next step had it not been for the overwhelming interest of his friends and students in the Cube. It was only after this unexpected success that even the thought of the possibility of producing it on an industrial scale came up. Eventually a manufacturer took on this job. Because of the inner complexity of the Cube, this was by no means straightforward and the process of turning Rubik's brainchild into thousands of low-cost, mass-manufactured units, progressed slowly and ultimately took some three years. The first Cubes subsequently appeared on the shelves of Budapest toyshops towards the end of 1977. It was not until January/February 1980, however, that the Cube made its international debut at the Toy Fairs of London, Paris, Nuremberg and New York, and export of the product began in May 1980. In 1981 demand exploded exponentially, from 1 million to 2 million, to 5 million, and production centers had to expand from Hungary to Brazil, Costa Rica, Hong Kong, and Taiwan. Hence, the Cube became a staggering commercial success and though it is difficult to estimate the total number of Cubes sold throughout the world, it is safe to assume that by 1998 the figure far exceeded 100 million. From this example, it is obvious that the unit cost of the very first Rubik's cube was high. Not only did Rubik himself spend several years engineering the prototype, also creating the ability to mass-produce the Cube took several years. After these relatively high development costs, the costs of actually producing a single Cube, the marginal costs, were small. They were just the cost of capital, labor and materials input, needed to produce it. As a consequence, the average cost of producing a Cube decreased: increasing returns to scale.

Another important property of knowledge is that some knowledge can only be created by men (and women; see footnote 2) of exceptional skill and ability. Even when given unlimited time, not everybody would be able to construct the theory of relativity as Einstein did. The law of large numbers thus plays a role in the creation of knowledge. The chance of finding a person who actually does have the exceptional skill and ability needed to create some knowledge increases with the number of people.

[4] The information on Rubik and his Cube is taken from Rubik's Online Homepage at http://www.rubiks.com/cubehistory.html (d.d. December 2003).

2.2 Empirical evidence

In early models of endogenous R&D-based growth, e.g. Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), the two above-mentioned properties of knowledge lead to a scale effect in growth: the size of the economy influences its growth rate. This survey of the empirical evidence starts with discussing suitable measures of size and empirical facts about GDP and population (which as the next paragraph will argue are important as measures of size). In the subsequent actual discussion of the empirical evidence, special attention is given to the seminal contribution of Jones (1995a) after which other available evidence is surveyed, namely Romer (1986), Backus, Kehoe and Kehoe (1992), Kremer (1993), Barro and Sala-i-Martin (1995), Frantzen (2002), and Papageorgiou (2002). At the end of the section, some concluding remarks are presented.

2.2.1 Suitable measures of size

For empirical research, it is of course extremely important to define the measure of size suitably (Backus, Kehoe and Kehoe 1992). For this purpose, the main mechanism driving the results of the models developed by Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), is investigated further. Refraining from microfoundations the essence of these models can be captured in three simple equations (see Jones 1999). Output (Y) is produced using a fraction of the total labor force (L_Y) and the stock of knowledge (A)[5]:

$$Y = A^{\sigma}L_Y \tag{2.1}$$

This production function exhibits constant returns to scale in the rivalrous inputs, labor, and increasing returns to scale in labor and the nonrivalrous inputs, knowledge, combined. That is, with $Y(A,L_Y)$ denoting the production process, the replication argument implies that $Y(A,\kappa L_Y) = \kappa Y(A,L_Y)$ (cf. Romer 1990, p. S76). Hence, a doubling of the output ($\kappa = 2$) is achieved by simply doubling the rivalrous inputs and no doubling of the nonrivalrous is required. The extent of these increasing returns is determined by the parameter $\sigma > 0$, so that

[5] Steedman (2001) calls the idea of a knowledge stock nonsense unless this stock is cardinally measurable. To address critique along these lines, far-fetched as it may be, beforehand, it is explicitly assumed here that the knowledge stock is cardinally measurable, thus giving 'meaning' to several of the equations used throughout this thesis. Chapter 5 elaborates on some of the problems related to knowledge and the way it is used in models of R&D-based growth.

in the situation depicted in equation (2.1), a doubling of all inputs, both rivalrous and nonrivalrous, would lead to total output being increased $2^{\sigma+1}$ times (> 2 times). New knowledge is created using a fraction of the total labor force (L_A) and the existing stock of knowledge:

$$\dot{A} = \delta L_A A \tag{2.2}$$

where δ is a parameter determining the arrival rate of new knowledge. From equation (2.2), it is clear that there is a knowledge spillover - and consequently a strong intertemporal R&D spillover - in the production of new knowledge, i.e. in the R&D process. The higher the current knowledge stock is, i.e. the more R&D has been performed in the past, the more productive the current R&D efforts are, and the faster the knowledge stock increases. That existing knowledge and/or past R&D play a role in the creation of new knowledge, partly determining the productivity of labor in the R&D process, is caused by the nonrivalrous nature of knowledge and ultimately leads to the prediction of a scale effect in growth. This is clarified using equation (2.3), which depicts the growth rate of output per worker (g_y):

$$g_y \equiv \frac{\dot{Y}}{Y} - \frac{\dot{L}}{L} = \sigma \frac{\dot{A}}{A} = \sigma g_A = \sigma \delta s L \tag{2.3}$$

where $0 < s < 1$ is the fraction of the total labor force active in the creation of new knowledge ($L_A = sL$; $L_Y = (1-s)L$). This last equation shows that in the early models of R&D-based endogenous growth, a permanent change in research intensity has a permanent effect on the growth rate. What is more, with population (L) increasing (over time), the growth rate of output per worker (per-capita income) increases (over time) as well: there is a scale effect in growth.

In addition, equation (2.3) presents a suitable measure of size (cf. Backus, Kehoe and Kehoe): an economy is larger than another economy if it devotes a larger absolute amount of resources, e.g. labor, to the production of knowledge (R&D) than the other economy, ceteris paribus. Thus if two countries spend a similar percentage of GDP (Gross Domestic Product) on R&D, the country with the highest GDP is also the largest. In addition, if two countries have a workforce (population) of equal size and productivity, the country that has the most people engaged in R&D activities, is the largest.

It is also often argued that a larger economy, be it measured by GDP or by population, provides a larger market for newly created knowledge, raising the return to R&D, thus stimulating R&D and ultimately raising the growth rate. Following this line of reasoning, any

variable that stimulates R&D, e.g. an R&D tax credit, can be used as a measure of size. However, since any stimulating effect on R&D will eventually show up in aggregate data on the absolute amount of resources a country devotes to the production of knowledge, the extent of R&D-stimulating incentives need not be incorporated explicitly in the measure of 'size'.[6]

2.2.2 Facts on GDP and population

Before starting the examination of empirical evidence on scale effects in growth, it seems wise to state facts regarding some variables important in the investigation of empirical evidence on scale effects in growth. There should be no ambiguity about the trends in both per-capita GDP and population. Through history, countries have generally experienced both an increase in per-capita GDP and an increase in population. This is further clarified in Figure 1 and Figure 2, which graph the natural logarithm of GDP per capita and the population respectively for France, Germany, Japan and the U.S., for the period 1870-1987.[7]

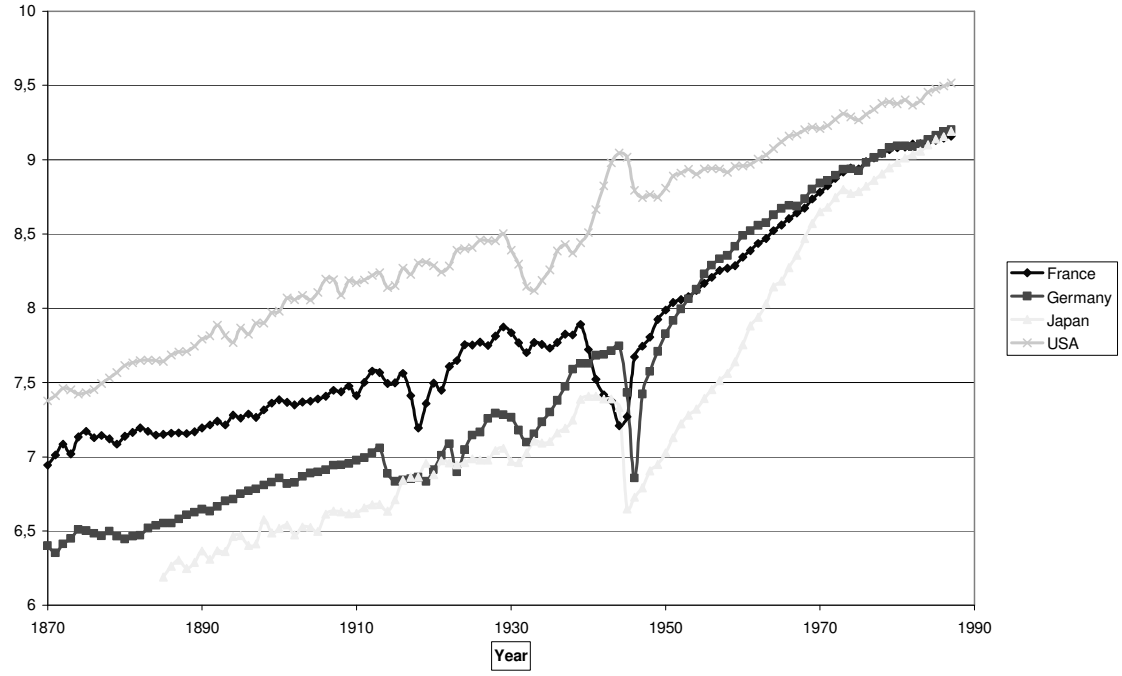


Figure 1, GDP per Capita 1870-1987.[8]

[6] Interesting studies following this line of reasoning are, for example, Sokoloff (1988) who finds a strong association between patenting and proximity to navigable waterways, and Ades and Glaeser (1999) who find that overall demand is important for growth and that openness is a substitute for initial wealth. These results support the idea that a large potential market stimulates innovation, i.e. the conduct of R&D.

[7] The graphs are limited to France, Germany, Japan, and the U.S., because these four countries are the main objects of interest in Jones' (1995a) influential paper on growth scale effects.

[8] Source: Jones (1995a) available at http://emlab.berkeley.edu/users/chad/TimeEGM.asc (d.d. December 2003).

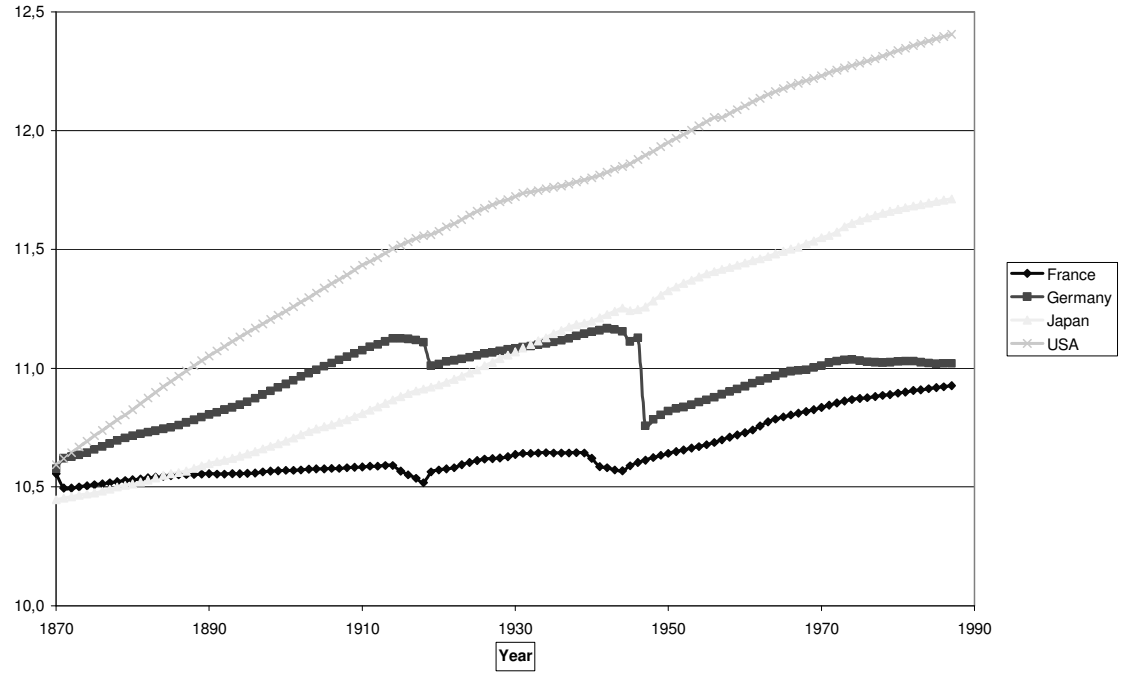


Figure 2, Population 1870-1987.[9]

Figure 1 clearly shows that GDP per capita for the advanced economies (France, Germany, Japan, and the U.S.) has followed a relatively steady path. Large fluctuations exist, but they generally correspond to external events, namely World War I, the Great Depression, and World War II. Figure 2 leads to the same conclusion as Figure 1; again, large fluctuations that generally correspond to external events are apparent.

2.2.3 The Jones evidence

As mentioned, Jones (1995a) put scale effects in growth on the map. In this paper Jones examines data for the U.S. and fourteen additional advanced OECD countries and finds that the growth rates of per-capita GDP for these countries have been remarkable stable, especially since the 1950s, and do not show a persistent increase (see also Figure 1).[10] This is inconsistent with the prediction of the early models of R&D-based growth (e.g. Romer, Grossman and Helpman, and Aghion and Howitt) that growth rates exhibit a permanent

[9] Source: Maddison (1991).
[10] Jones (1995a) uses several tests to determine whether the growth rates of per-capita GDP of these countries are indeed stable over time and finds for instance that an augmented Dickey-Fuller test strongly rejects the null hypothesis of a unit root in growth rates over the period 1900-1987. Jones (1995a) also derives point estimates of time trend coefficients for total factor productivity (TFP) growth for France, Germany, Japan, and the U.S. These estimates are uniformly negative, though only significantly so for France and Japan.

increase. Jones states two possible conclusions that might be drawn from this empirical evidence.

The first is that (1) the movements of the variables that might have an impact on growth rates of per-capita GDP cancel each other out. Over the last forty years, however, a lot of variables that might positively influence the growth rate, have shown a persistent increase. Among these variables are investment (as a share of GDP), average years of education per adult, and the number of people active in R&D. Figure 3 depicts the natural logarithm of the number of scientists and engineers active in R&D for France, Germany, Japan, and the U.S., for the period 1950-1987 (1965-1987 respectively).

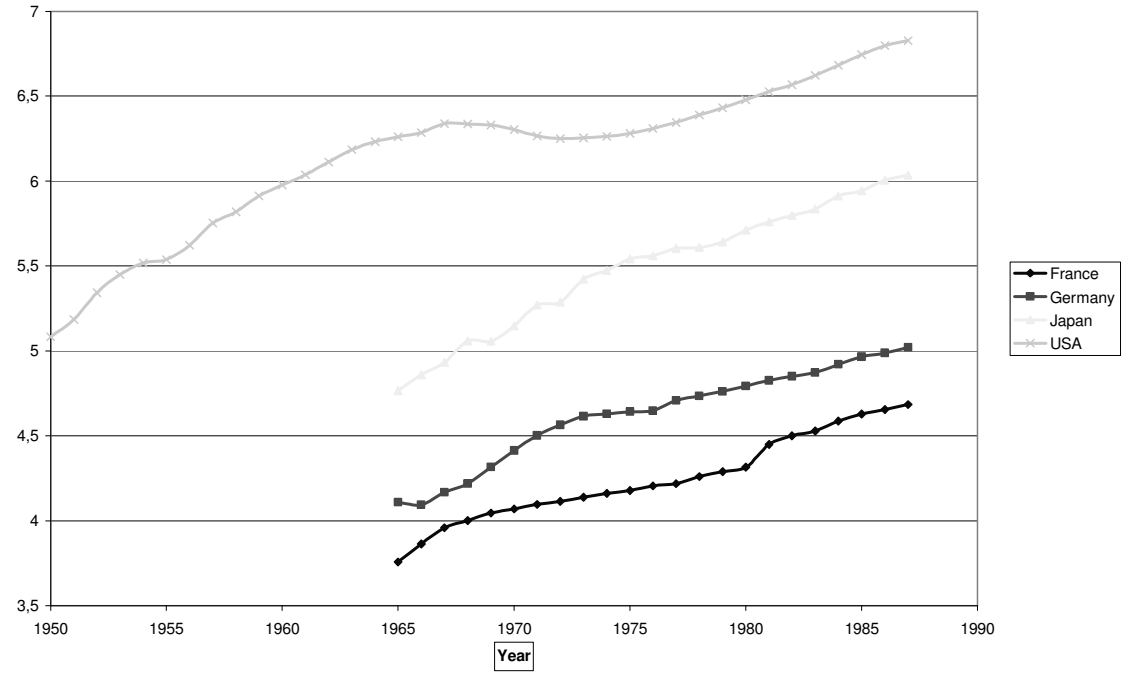


Figure 3, Scientists and engineers engaged in R&D.[11]

Figure 3 strengthens the idea that it is very unlikely that this first conclusion is correct. The second far less unlikely conclusion is that (2) the prediction of several early models of R&D-based growth that permanent changes in some variables have permanent effects on growth rates of per-capita GDP is simply false.

Jones' findings are probably best summarized in the restriction he places on models of endogenous (R&D-based) growth: "… if an endogenous growth model predicts that permanent movements in some variable X have permanent effects on growth, then either

[11] Source: Jones (1995a) available at http://emlab.berkeley.edu/users/chad/TimeEGM.asc (d.d. December 2003).

a) X must exhibit no large persistent movements, or
b) some other variable (or variables) must also have persistent effects on growth that offset the movements of X in a way that is determined by the endogenous growth model." (p. 502).

2.2.4 Other evidence

Jones (1995a) has started a lot of work on scale effects on growth, both empirical and theoretical (see Section 2.3). Some other notable contributions to empirical evidence on scale effects in growth are, amongst others, Romer (1986), Backus, Kehoe and Kehoe (1992), Kremer (1993), Barro and Sala-i-Martin (1995), Frantzen (2002), and Papageorgiou (2002).

Romer (1986)

This study sparked the development of what is now often referred to as new growth theory (or endogenous growth theory; see also Lucas 1988). In Romer's (1986) model of endogenous growth, the rate of growth can be increasing over time and large countries grow faster than small countries ceteris paribus. To find evidence for this scale effect prediction regarding growth rates of his model, he compares growth rates of per-capita income over time. He finds that in a sample of 11 OECD countries for which continuous observations on GDP per capita are available starting no later than 1870 per-capita income growth increases across decades. To compare the growth rates of different decades, Romer uses a statistic (π) that gives the sample estimate of the probability that, for any two randomly chosen decades, the later decade has a higher growth rate. Over the entire sample of 11 countries, the estimated value for π ranges from 0.58 for Sweden to 0.81 for Norway. The corresponding p-values of observing a value of π at least as large as the observed value under the null hypothesis that the true probability is 0.5 range from 0.25 for Sweden to 0.002 for Norway. For 5 out of these 11 countries, the null hypothesis that there is no positive trend in the growth rate is rejected at the five percent significance level. For 8 out of these 11 countries, the null hypothesis is rejected at the ten percent significance level. This observed increase of growth over decades may of course be evidence of a scale effect in growth since over the decades also GDP, population, and consequently resources used in the production of knowledge (R&D), have increased.

Backus, Kehoe and Kehoe (1992)

In their study, Backus, Kehoe and Kehoe relate differences in growth rates to differences in the scale of production and inputs. They find no statistically significant evidence for the existence of a scale effect in a cross-sectional regression of the growth rate of GDP per capita on several measures of scale (size). Their study does, however, find highly statistically significant evidence on the existence of a growth scale effect within a single sector, namely manufacturing. Growth in manufacturing output per worker increases with the natural logarithm of initial manufacturing output and other measures of scale such as the (natural logarithm of) the total number of scientists and total R&D expenditure.

Kremer (1993)

Kremer constructs a model that combines two assumptions: (1) that due to the nonrivalrous nature of technology the growth rate of technology is proportional to the level of population, and (2) that the growth rate of (world) population is proportional to the growth rate of technology. The model thus predicts that, as a result of the nonrivalrousness of technology (knowledge), the growth rate of world population is proportional to its level. This prediction is corroborated by empirical evidence: through most of history, the growth rate of world population has been approximately proportional to the level of population the implication for growth of course being that it exhibits a scale effect.

Barro and Sala-i-Martin (1995)

To test for some sort of benefit from a larger scale, Barro and Sala-i-Martin include the natural logarithm of the working-age population in their regression analysis of a cross-section of countries. The estimated coefficient for this variable has a positive sign but is not significant at usual critical levels.

Frantzen (2002)

The research done by Frantzen is very interesting in that it tries to combine the issue of scale effects in growth with the extent of R&D (knowledge) spillovers. The importance of knowledge spillovers for the growth of per-capita income and/or the growth of productivity has been shown empirically.[12] In the knowledge production function depicted in equation

[12] See for instance Coe and Helpman (1995) and Coe, Helpman and Hoffmaister (1997) for evidence on international spillovers, and Verspagen (1997) for evidence on both international and intersectoral (domestic) spillovers.

(2.2), the extent of knowledge spillovers is such that they give rise to a scale effect in growth. Frantzen examines knowledge spillovers and scale effects in growth in the context of a Cobb-Douglas type of aggregate production function that exhibits constant returns to scale in the rivalrous inputs, capital and human capital, and increasing returns to scale to a measure of total factor productivity (TFP) and the rivalrous factors combined.[13] This production function underlies his test of data on 14 OECD countries for the existence of two possible situations.

1. In the first, the strength of the knowledge spillovers is not large enough to allow for a scale effect in growth, and a direct (linear) relation between the level of TFP and the level of R&D capital (knowledge) is expected.
2. In the second possible situation, the knowledge spillovers are strong enough to lead to a scale effect in growth and a direct (linear) relation between the growth rate of TFP and the level of R&D capital is expected.

In this setting spillovers are allowed for by constructing TFP to be a function of four distinct forms of R&D capital, (1) own sector domestic, (2) other sector domestic, (3) same sector foreign, and (4) other sector foreign R&D capital. The framework thus explicitly takes into account the effects of domestic intersectoral, international intrasectoral, and international intersectoral spillovers. The contribution of the three types of R&D capital associated with these spillovers (i.e. other sector domestic, same sector foreign and other sector foreign R&D capital, but not own sector domestic R&D capital) to (domestic) TFP is weighted by the technological distance between the senders and receivers measured using a patent-based method (see also Verspagen 1997). The use of a measure of R&D capital corrected for technological distance reveals that the study of Frantzen investigates the extent of pure knowledge spillovers as opposed to the studies by Coe and Helpman (1995) and Coe, Helpman and Hoffmaister (1997) mentioned in footnote 12, which study the extent of so-called rent spillovers. The difference between these two types of spillovers is that pure knowledge spillovers are directly related to knowledge's nonrivalrous nature, while rent spillovers result from R&D intensive intermediate goods which are bought at a price that does not fully reflect their quality (consumer surplus). Coe and Helpman (1995) and Coe, Helpman and Hoffmaister (1997) use R&D capital that is weighted by trade shares, which makes their framework more suited to study the extent of rent spillovers. In addition, these studies are

[13] A simple definition would be that TFP growth is growth of GDP left unexplained when increases in the other inputs, capital and human capital, are accounted for. There is a close link between knowledge and TFP as is apparent when this definition is compared with the preliminary definition of knowledge given above. See also Chapter 5.

limited to international spillovers and do not take into account the separate effect of intersectoral spillovers, both domestic and foreign, on TFP. Because, in light of the subject of this thesis, it is important to distinguish between the two types of spillovers, only pure knowledge spillovers are of interest (see also the next chapter), Appendix A provides an extended discussion of the two different types of spillovers and the frameworks used to study their effect on for example TFP. Using much the same test methods as Jones (1995a) (e.g. unit-root tests), Frantzen subsequently rejects the second possibility of the existence of a scale effect in growth. However, he also finds that the elasticity of TFP with respect to the different types of R&D capital (own sector domestic, other sector domestic, same sector foreign, and other sector foreign) is substantial and statistically significant for each type. Hence, knowledge spillovers are very much present just not to the extent that they lead to a scale effect in growth.

Papageorgiou (2002)

Papageorgiou presents some cross-country and time-series evidence that points to the existence of growth scale effects in the early stages of development. He performs a cross-country regression in which growth of per-capita GDP is modelled as a function of the level of human capital, as measured by either an index of the average years of education of the country's citizens or by the number of scientists and engineers in the country. Papageorgiou finds a hump-shaped (inverse U-shaped) relation: the growth rate of GDP per capita increases with the level of human capital until it reaches a certain level after which it decreases with the level of human capital. To further investigate this relation, he divides the countries in his data set over four income classes, low-income (less than $2000), low/middle-income ($2001-$5500), middle-income ($5501-$9000), and high income (more than $9001). Regression analysis subsequently shows a statistically significant positive linear relation between the growth rate of per-capita income and the level of human capital for both the low- and low-middle income countries whilst such a relationship is absent (i.e. not significant) for the middle- and high-income countries. This supports the idea of a hump-shaped relation between the level of human capital and per-capita GDP growth. A final piece of evidence for the hump-shaped hypothesis is found in a segmented time-trend test for 15 OECD countries. Papageorgiou examines two epochs, 1871-1949 and 1950-1987, and finds that most of the countries examined experienced statistically significant growth rates of per-capita GDP in the first epoch, while the calculated growth rates for the second epoch are generally insignificant and/or lower.

2.2.5 Concluding remarks

The papers discussed above present mixed evidence on scale effects in growth. Furthermore, some important critical remarks regarding (some of the) papers and their methodology must be made.

1. Country differences might not be found because knowledge is diffused across countries rather fast. This means that not a country's level of GDP or population is the suitable measure of size, but that instead world income and world population should be used.[14] Another, perhaps even more suitable measure of size would be the amount of money spend on R&D (in a country or in the world as a whole) which in turn is to no small extent determined by other measures like total market size, e.g. GDP (see also Sokoloff 1988 and Ades and Glaeser 1999).
2. Accelerating growth in GDP per capita is only prima facie evidence on a growth scale effect. That is to say, any observed acceleration of growth might be due to the acceleration of growth of productive inputs, namely capital and human capital, where human capital is defined as the number of effective units of labor (units of labor corrected for quality as measured by years of education for instance). The relation between knowledge, which unique properties lay at the root of scale effects in growth, and growth in per-capita GDP in general, is thus not one on one. Evidence on the existence of a growth scale effects should not be sought in (accelerating) growth of GDP per capita, but rather in the behavior over time of another variable that has a closer relation to knowledge, e.g. TFP. As noted in footnote 13, TFP is closely related to knowledge (see also Chapter 5) and consequently is the most logical variable to look at when investigating a possible scale effect in growth. Undoubtedly, an accelerating rate of growth in TFP is more convincing evidence on scale effects, than is accelerating growth of per-capita GDP.
3. Accelerating growth may simply not show itself in an increase in GDP per capita. The idea is that technological progress (i.e. the creation of new knowledge) leads to quality improvements and the (faster) introduction of new products and consequently contributes to growth in per-capita income in a way that (conventional) measures of prices (and GDP)

[14] Similarly, Jones (1995b) notes that "Comparing the U.S. and its population growth rate to Japan and its population growth rate would neglect the potentially large flow of ideas that occurs between these countries. The world (or at least that part of the world that is growing by pushing out the technological frontier) is potentially a more appropriate unit of observation." (p. 777/778). However, studies like those by Frantzen and Papageorgiou have rejected the existence of a scale effect in growth in a data set containing countries which account for the bulk of worldwide production and R&D effort.

cannot account for accurately (see Chapter 5 for a further discussion of this measurement problem).

Bearing in mind these critical remarks, drawing a definite conclusion about the existence of scale effects in growth might seem unwise. Indeed, future research on scale effects in growth should address the above-made critical remarks. These critical remarks notwithstanding, the evidence against scale effects in growth presented in Jones (1995a) (time-series evidence) and to a lesser extent Backus, Kehoe and Kehoe (cross-sectional evidence) remains compelling, and hence the conclusion that growth does not exhibit a scale effect is justified. This is not to say, however, that (i) knowledge is a rivalrous good rather than a nonrivalrous good, (ii) that the nonrivalrous nature of knowledge does not lead to increasing returns, and/or (iii) that market size does not influence the amount of R&D performed (see for instance Sokoloff and Ades and Glaeser). In line with Jones' (1995a) conclusion regarding scale effects and models of growth, the conclusion here is simply that there are no scale effects in growth and that (realistic) R&D-based models of growth should incorporate a mechanism that mitigates possible scale effects in growth arising from, for instance, the nonrivalrousness of knowledge.

2.3 Models of growth without scale effects

Jones (1995a) spurred the development of R&D-based models of growth without scale effects. Notable contributions are, amongst others, Jones (1995b), Smulders and van de Klundert (1995), Kortum (1997), Aghion and Howitt (1998), Dinopoulos and Thompson (1998), Peretto (1998), Segerstrom (1998), Young (1998), Li (2000 and 2002), and Peretto and Smulders (2002).

This discussion follows Jones (1999) and his division of models of growth without scale effects in two broad categories and adds a category for recent models that do not feature in his taxonomy, e.g. Li (2000 and 2002) and Peretto and Smulders (2002). The section ends with some concluding remarks.

2.3.1 One-R&D-sector semi-endogenous growth models

Jones' (1999) first category contains models that can best be described as one-R&D-sector semi-endogenous models of growth without scale effects (cf. Li). The models that he shares under this category are Jones (1995b), Kortum (1997), and Segerstrom (1998). These models

are one-dimensional in the sense that R&D leads to either quality improvement or variety expansion, and semi-endogenous in the sense that the creation of knowledge requires real resources (endogenous), but that the long-run growth rate cannot be affected by government actions, e.g. an R&D tax credit (exogenous).

Jones (1995b), Kortum (1997) and Segerstrom (1998), generalize the R&D production function used in Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992). Refraining from microfoundations, the essence of their models can be captured in three equations (see Jones 1999). The knowledge production function is very similar to that presented in equation (2.2) and is given by:

$$\dot{A} = \delta L_A A^{\phi} \tag{2.4}$$

where ϕ measures the degree of externalities in the R&D process across time, i.e. the extent of intertemporal R&D spillovers. The situation of $\phi < 0$, corresponds to 'fishing out' (Jones 1998a): a situation in which the most obvious knowledge is created first and subsequent new knowledge is increasingly difficult to create. If $\phi > 0$, existing knowledge increases future R&D productivity, a situation that Jones (1998a) refers to as 'standing on shoulders' (see also Caballero and Jaffe 1993). With $\phi = 0$, the existing stock of knowledge has no influence on the creation of new knowledge; there are no intertemporal externalities in the R&D process. Finally, if $\phi = 1$, equation (2.4) resembles the knowledge production function of Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992) (see equation 2.2).

Jones (1995b)

Jones uses the generalized knowledge production function depicted above to argue that the assumption of $\phi = 1$ is arbitrary (see also Frantzen 2002; cf. Solow 1994). He subsequently imposes the restriction of $\phi < 1$ to create a model in which along the balanced growth path the number of persons engaged in R&D increases. Underlying the assumption of $\phi < 1$ is the idea that the amount of labor required to create new knowledge increases with the level of knowledge already obtained. Since along the balanced growth path, the growth rate of knowledge is constant by definition[15], this growth rate (g_A) can easily be calculated from equation (2.4):

[15] Along a balanced growth path the derivative of a certain growth rate with respect to time is zero.

$$g_A = \frac{n}{1-\phi} \tag{2.5}$$

where n is the growth rate of the total labor force (population growth). Combining equation (2.3) and (2.5) yields the growth rate of output per worker (g_y):

$$g_y = \frac{\sigma n}{1-\phi} \tag{2.6}$$

Equation (2.6) clearly shows that in Jones' model, the growth rate does not depend on size (as measured for instance by total population), but depends only on the (exogenous) growth rate of the labor force and the value of the parameters ϕ and σ. In fact, as mentioned above, the model is semi-endogenous in the sense that the creation of knowledge requires real resources but that long-run growth cannot be affected by government actions. Another important feature is that although the model has no scale effect in growth, the model does exhibit a level scale effect, i.e. per-capita income is higher in larger economies (with of course size suitably measured). The existence of a level scale effect follows, with a bit of rearranging, simply from the above equations. Using equation (2.1), per-capita income (y) can be written as:

$$y = \frac{Y}{L} = A^{\sigma}(1-s) \tag{2.7}$$

The knowledge stock can be specified further by combining equation (2.4) and (2.5) which yields the following equality:

$$g_A = \frac{\dot{A}}{A} = \delta L_A A^{\phi-1} = \delta s L A^{\phi-1} = \frac{n}{1-\phi} \tag{2.8}$$

The knowledge stock can subsequently be rewritten as a function of time:

$$A(t) = \left[\frac{n}{(1-\phi)\delta s L}\right]^{\frac{1}{\phi-1}} = \left[\frac{(1-\phi)\delta}{n} sL(t)\right]^{\frac{1}{1-\phi}} \tag{2.9}$$

This ultimately renders an expression for the development of per-capita income (output) over time:

$$y(t) = (1-s)\left[\frac{(1-\phi)\delta}{n} sL(t)\right]^{\frac{\sigma}{1-\phi}} \tag{2.10}$$

It is clear from this last equation that (changes in) size or changes in R&D intensity (s) do effect the long-run (steady state) level of per-capita income, i.e. there is a level scale effect[16]. This level scale effect reflects the idea that a larger economy, by providing a larger market for newly created knowledge, stimulates R&D, ultimately raising the growth rate (see also Sokoloff and Ades and Glaeser). In addition, as opposed to the long-run growth rate of per-capita income, the long-run level of per-capita income can be influenced by government action, for example through the careful use of R&D subsidies/taxes to induce the R&D intensity ($s = \sigma/(1+\sigma-\phi)$) that maximizes the level of per-capita income.[17] Note however, that there is no evidence supporting the existence of a level scale effect. For instance, the United States, Japan and Germany are not richer or more productive than Belgium, Luxembourg and Switzerland. The reason behind this is probably the large flow of knowledge (through spillovers) between these countries and countries in general. Such a knowledge flow limits potentially large differences in productivity between countries with similar capital and human capital endowments (see also footnote 14).

Kortum (1997)

In his model, Kortum provides an elegant underpinning for Jones' (1995b) fishing-out effect ($\phi < 1$) that does not rely on the most obvious ideas being discovered first. Kortum tries to explain the long-run behavior of R&D employment, patenting, and growth of total factor productivity. Since the 1950s, research employment has increased, the number of patents per researcher has declined, and growth of TFP has not increased with the level of R&D. In order to explain these puzzling trends, he develops a search-theoretic model in which R&D is a process of search that draws samples from probability distributions of potential new production techniques. The current state of the art production technique is the result of past R&D and the more the state of the art production technique advances, the more difficult it becomes to draw an even more efficient production technique from the probability distributions of potential new production techniques. The R&D process is thus very much like throwing a dice (twice). If you have thrown 1 in your first throw the chance that you will

[16] The scale effects that are of interest in this thesis are, however, growth scale effects and not level scale effects. Any reference to scale effects or a scale effect is consequently to the growth variant and not to the level variant. What is more, in the event that level scale effects are the type of scale effects that actually are of interest, it will be made clear that the concern is with this type and not with the growth type.

[17] An increase in the R&D intensity of course also temporarily raises the growth rate beyond the rate associated with the balanced growth path and hence the growth rate can be affected (endogenous) through policy. The growth effect of for instance a policy-induced increase in R&D intensity is, however, temporary and only lasts until the steady state (balanced growth path) is reached again.

throw a higher number in your second throw is relatively high (5/6). On the other hand, if you have thrown 5 in your first throw, the chance that you will throw a higher number in your second throw is relatively low (1/6). Kortum uses this microfoundation of why R&D becomes increasingly difficult to explain why patenting has been roughly constant while the number of persons active in R&D has risen sharply since the early 1950s. In addition, in his model the rate of productivity growth is proportional to the growth rate of the research stock (i.e. the progressing of the state of the art production technique)[18] which corresponds to the fact that over the last four decades or so, the growth rate of TFP has not risen with the level of R&D. Thus the model does not exhibit a growth scale effect though, similar to Jones (1995b), it does exhibit a level scale effect. In Kortum's model R&D employment (and ultimately economic growth) is determined by the value of patents and the probability of actually discovering a new state of the art production technique relative to the wage of labor used in producing output with the current state of the art production technique. R&D employment subsequently increases as the value of patents relative to wages increases, and this value increases with a growing labor force (population growth).

Segerstrom (1998)

The model developed by Segerstrom is similar to that of Jones (1995b). Like Kortum's model, however, Segerstrom's model is not only able to explain the absence of scale effects in growth (though similar to Jones 1995b and Kortum 1997 it does exhibit a level scale effect), but also the relatively constant rate of patenting over the last few decades[19].

2.3.2 Two-R&D-sector endogenous growth models

Jones' (1999) second category contains models which can best be described as two-R&D-sector endogenous models of growth without scale effects (cf. Li). The models that Jones (1999) shares under this category are Aghion and Howitt (1998), Dinopoulos and Thompson (1998), Peretto (1998), and Young (1998). In contrast with the models by Jones (1995b), Kortum (1997) and Segerstrom (1998), these models are two-dimensional in the sense that they allow for both (1) R&D that leads to quality enhancements and for (2) R&D that

[18] That the rate of productivity growth is proportional to the growth rate of the research stock follows from Kortum's assumption that the search distribution of potential new techniques is Pareto.

[19] As Segerstrom notes, the number of patents granted to residents of the USA in 1993 (53,200) is almost equal to the number in 1966 (54,600). This observation in turn is related to Kortum's observation that research employment has increased and the number of patents per researcher has declined since the 1950s.

increases variety, and endogenous in the sense that the creation of knowledge requires real resources and that the long-run growth rate can be affected by government actions. The models of R&D-based growth by Aghion and Howitt (1998), Dinopoulos and Thompson (1998), Peretto (1998) and Young (1998), thus not only relieve the scale effect property, they also maintain the endogeneity of long-run growth.

Young is one of the firsts to develop a two-R&D-sector model of endogenous growth.[20] His model incorporates the concept of 'equivalent innovation' which says that an increase in the market size and the correspondingly increasing profitability of inventing a solution to a problem leads to a greater variety of solutions that solve the same problem.[21] This increase in variety raises the average level of consumer utility. Further improvement of this greater variety of solutions requires additional R&D however. Consequently the equilibrium level of expenditure on R&D may rise without this rise leading to a corresponding increase in the growth rate of product-quality improvement, and hence the rate of per-capita income growth.[22] Young subsequently combines this concept with some of the standard features of the early models of R&D-based growth, e.g. the nonrivalrous nature of knowledge and the proportionality of spillovers ($\phi = 1$ in equation 2.4; see also equation 2.2). Refraining from microfoundations, the essence of the model of Young (1998) and the other above-mentioned two-R&D-sector endogenous growth models, namely Aghion and Howitt (1998), Dinopoulos and Thompson (1998) and Peretto (1998), is captured by some simple equations that add to the core of the early endogenous growth models developed in equations (2.1) through (2.3) (see Jones 1999).[23]

Aggregate consumption and/or output (C) is given by the familiar Dixit-Stiglitz production function (Dixit and Stiglitz 1977):

$$C = \left[\int_0^B Y_i^{1/\theta} di \right]^{\theta} \tag{2.11}$$

[20] Young readily acknowledges that other researchers, e.g. Smulders and van de Klundert (1995), pioneered the development of two-R&D-sector endogenous growth models. The contribution of Smulders and van de Klundert is, however, not included in Jones' (1999) discussion of this type of R&D-based models of growth without scale effects.

[21] Young speaks of solutions to problems. In the terminology used throughout this thesis, these solutions would be referred to as (new) knowledge.

[22] In the pioneering study of Smulders and van de Klundert, the same effect is reached by assuming that the core of knowledge is tied to an organization in tacit form, and thus is firm specific (see also Dosi 1988).

[23] This generalization of course does not do justice to the richness of each of these models, their assumptions, and their microeconomic foundations. It does provide, however, an easy understanding of the working of two-R&D-sector models of endogenous growth without scale effects, including the early one by Smulders and van de Klundert.

where Y_i is the consumption of variety i, B measures the number of varieties available for consumption, and $\theta > 1$ is related to the elasticity of substitution between different varieties. The production function for varieties is similar to the knowledge production function depicted in equation (2.2):

$$\dot{B} = LB^{\gamma} \tag{2.12}$$

where γ measures the degree of externalities in the variety creating process across time (intertemporal variety spillovers). Along the balanced growth path the total variety of consumption goods simply evolves over time according to:

$$B = L^{\beta} \tag{2.13}$$

where $\beta = 1/(1-\gamma)$ is any real number. To further solve the model, it is assumed that each intermediate good Y_i is used in the same amount so that $Y_i = Y$ and $C = B^{\theta}Y$. Output per capita (c) is then given by:

$$\frac{C}{L} = c = B^{\theta}\frac{Y}{L} = B^{\theta}y \tag{2.14}$$

Growth of per-capita output (g_c) subsequently equals:

$$g_c = \theta g_B + g_y = \theta\beta n + \sigma g_A \tag{2.15}$$

where n (again) depicts the growth of the labor force. Young's concept of equivalent innovation means that the knowledge production function of the early models of R&D-based growth (see equation 2.2), modifies to:

$$\dot{A} = \delta A\frac{L_A}{B} \tag{2.16}$$

It is thus no longer the absolute amount of labor active in R&D that matters in the production of knowledge, but the (relative) amount of R&D labor per variety. Hence, the concept of equivalent innovation in principal limits the extent of knowledge's nonrivalrousness. The same result is more straightforwardly achieved by assuming that knowledge is firm specific (cf. Smulders and van de Klundert). The growth rate of knowledge (g_A) subsequently rewrites to:

$$g_A = \delta s\frac{L}{B} = \delta s L^{1-\beta} \tag{2.17}$$

Finally, the growth rate of per-capita output equals:

$$g_c = \theta\beta n + \sigma\delta s L^{1-\beta} \tag{2.18}$$

The key results of the models of Young (1998), Aghion and Howitt (1998), Dinopoulos and Thompson (1998) and Peretto (1998), that growth does not exhibit a scale effect but is endogenous, is reached by implementing the condition that $\beta = 1$. In this case, the growth rate does not depend on a measure of size (i.e. L), while changes in the fraction of the total labor force active in R&D (s) do affect the growth rate. Similar to the assumption that $\phi = 1$ in equation (2.4), the assumption that $\beta = 1$ is arbitrary however (cf. Solow 1994). In fact, in the case of $\beta = 1$ it is simply assumed that there are no externalities (spillovers) in the variety creating process ($1/(1-\gamma) = 1$; $\gamma = 0$; $\dot{B} = L$). What is more, similar to the one-R&D-sector models of growth without scale effects, these models exhibit a level scale effect; combining equation (2.13) and (2.14) straightforwardly shows, that output/consumption per capita (c) increases with size.

2.3.3 Recent models of growth without scale effects

Jones' (1999) discussion of the models of growth without scale effects developed by Jones (1995b), Kortum (1997), Segerstrom (1998), Young (1998), Aghion and Howitt (1998), Dinopoulos and Thompson (1998) and Peretto (1998), has stressed two important shortcomings:

1. in the models that incorporate only one R&D sector, long-run growth is semi-endogenous, i.e. long-run growth cannot be affected by policy actions[24],
2. the models that incorporate two R&D sectors, make an arbitrary assumption about spillovers to remove the scale effect.

Other recent models of growth without scale effects extend the analysis of the models discussed above and in doing so address their shortcomings (e.g. Li and Peretto and Smulders).

[24] Of course the exogeneity of long-run growth does not preclude the government's use of for instance R&D subsidies/taxes to influence the rate of per-capita income growth in the short-run (see also the discussion of level scale effects in Paragraph 2.3.1).

Li (2000 and 2002)

As opposed to the models discussed above, in Li's two-R&D-sector model of growth without scale effects spillovers between the two types of R&D (quality enhancing and variety increasing) do matter. Introducing quality-variety spillovers to these models generalizes equations (2.12) and (2.16) to:

$$\dot{A} = \delta s L A^{\phi_A} B^{\gamma_A} \tag{2.19}$$

$$\dot{B} = L B^{\gamma_B} A^{\phi_B} \tag{2.20}$$

where ϕ_A, γ_A, γ_B and ϕ_B, measure the degree of externalities over time in the product enhancing and variety creating process respectively (spillovers).[25] Along the balanced growth path the growth rates of A and B can simply be found to equal:

$$g_A = \frac{n + \gamma_A g_B}{1 - \phi_A} \tag{2.21}$$

$$g_B = \frac{n + \phi_B g_A}{1 - \gamma_B} \tag{2.22}$$

Using Cramer's rule, the system of equations depicted in equations (2.21) and (2.22) can be solved:

$$g_A = \frac{1 - \gamma_A - \gamma_B}{D} n \tag{2.23}$$

$$g_B = \frac{1 - \phi_A - \phi_B}{D} n \tag{2.24}$$

with $D = (1-\phi_A)(1-\gamma_B) - \gamma_A\phi_B \neq 0$ (D is the determinant of the coefficient matrix of the system of equations depicted in equations 2.21 and 2.22). Generalizing the knowledge production functions leads to some interesting conclusions.

1. If $D \neq 0$, the long-run rate of growth of income per capita ($g_c = \theta g_B + \sigma g_A$) is not endogenous and is determined by population growth (n) (exogenous) and the parameters that measure the extent of the knowledge spillovers (ϕ_A, γ_A, γ_B, and ϕ_B) (exogenous) as is immediately clear from equation (2.23) and (2.24). This is in contrast with other two-R&D-sector models of growth without scale effects in which long-run growth is endogenous (e.g. Young 1998, Aghion and Howitt 1998, Dinopoulos and Thompson

[25] Young's concept of equivalent innovation is here captured in the parameter γ_A. In the notation of Li, both equation (2.19) and (2.20) have an additional term, B^{-1} and A^{-1} respectively.

1998, and Peretto 1998). In Li's model, which allows for spillovers between the two different R&D sectors (i.e. quality and variety), endogenous long-run growth, that is a rate of growth that can be affected through policy measures, requires that D = 0 for only then the growth rates of A and B (and consequently c) do not (necessarily) depend on exogenous parameters (ϕ_A, γ_A, and so forth). This, however, requires satisfaction of two knife-edge conditions: (1) $\phi_A = \phi_B + 1$, and (2) $\gamma_A = \gamma_B - 1$.[26]

2. As the two knife-edge conditions are not easily satisfied, long-run growth in two-R&D-sector models (with quality-variety spillovers) is generally exogenous. In fact, the main result of Li is that in k-R&D-sector models, k knife-edge conditions need to be satisfied in order for growth to be endogenous. Hence, exogenous long-run growth resulting from the one-R&D-sector models developed by Jones (1995b), Kortum (1997) and Segerstrom (1998), is more general than endogenous long-run growth resulting from the two-R&D-sector models developed by Young (1998), Aghion and Howitt (1998), Dinopoulos and Thompson (1998), and Peretto (1998).[27] The driving force behind this result is Li's assumption of spillovers between the two different types of R&D, quality enhancing and variety increasing.

Peretto and Smulders (2002)

Peretto and Smulders provide the microfoundation for the particular spillover function that drive the results of the above-discussed two-R&D-sector models of growth. As mentioned, the result of these models, endogenous long-run growth without scale effects, hinges importantly on the assumption that $\beta = 1$ ($\gamma = 0$). Young does, however, examine the situations in which either $\beta < 1$, or $\beta > 1$. If $\beta < 1$, the two-R&D-sector models have the features of the early models of R&D-based growth, namely growth with scale effects (see Jones 1999). On the other hand, if $\beta > 1$, the results of these models are returned to those of one-R&D-sector models of growth, namely semi-endogenous growth without scale effects. In the model developed by Peretto and Smulders both the situation of $\beta < 1$ (positive scale effects) and that of $\beta > 1$ (negative scale effects) can exist, but both situations vanish asymptotically with the size of the economy. That is, in their model the situation in which $\beta =$

[26] In Li, the adding of the terms B^{-1} and A^{-1} to equation (2.19) and (2.20) respectively (see footnote 25 above), leads to a simpler formulation of these knife-edge conditions ($\phi \equiv \phi_A = \phi_B$ and $\gamma \equiv \gamma_A = \gamma_B$).

[27] In the models of Jones (1995b), Kortum (1997) and Segerstrom (1998), only one knife-edge condition ($\phi = 1$) needs to be satisfied to make growth endogenous. Making growth endogenous in their framework returns their models to those developed by Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), i.e. a model of growth with scale effects.

1, which seems to be a knife-edge situation, is always obtained as an economy increases in size. Hence, in their framework growth scale effects disappear almost naturally (with the economy increasing in size, which happens naturally over time). Peretto and Smulders allow for the accumulation of knowledge along two dimensions: (1) the intensive margin (R&D leads to improved understanding of existing problems) and the (2) extensive margin (original formulation of new problems). The type of accumulation, i.e. accumulation along the intensive margin or along the extensive margin, affects the extent of spillovers. Accumulation along the intensive margin gives rise to spillovers by increasing the public knowledge stock (positive scale effects). Accumulation along the extensive margin on the other hand leads to dilution of spillovers, because technological distance increases and the relevance of the public knowledge stock for a given firm's R&D becomes smaller (negative scale effects). Ultimately, however, as the size of the economy increases, both types of scale effects vanish asymptotically. The root for this interesting feature of the model developed by Peretto and Smulders lies in their assumption that the structure of an economy, is not symmetrical in size, i.e. the structure of a small economy is different than that of a large economy. This follows from the idea that an economy consists of different networks in which firms participate. These networks do not increase in size with the size of the aggregate economy (asymmetrical) but rather larger economies consist of more networks. This results from the fact that size attracts entrants who bring with them new products and consequently new networks centered around these new products. In this framework imposing a simple limit on the extent of the spillovers between the various networks, leads to growth that does not exhibit a scale effect ($\beta = 1$).[28] This limit in turn is due to technological distance between the different networks, which, as mentioned above, diminishes the relevance of spillovers.

2.3.4 Concluding remarks

Having discussed some important proposed theoretical solutions to the 'scale effects problem', some conclusions can be drawn. The three most important ones are that:

1. All (types of) models discussed are not without merit. They have provided two broad categories of 'solutions' to the scale effects 'problem'. The first category can best be referred to as (1) the fishing-out effect (cf. Jones 1998a), which basically states that with a

[28] The model of Peretto and Smulders does, however, exhibit a level scale effect, as does Li 's model and, in fact, all the models of growth without scale effects discussed in this section.

rising knowledge stock, the creation of new knowledge becomes increasingly more difficult. The second category can best be referred to as (2) the dilution effect (cf. Peretto and Smulders), which basically states that the larger the knowledge stock is, the larger also technological distance is, and the less useful this stock is for a given economic subject;

2. Almost all (types of) models make assumptions that can be criticized, especially the assumptions regarding knowledge and intertemporal R&D spillovers, which are either absent or severely limited, through the two knowledge externalities discussed above. This makes Solow's (1994) critique on R&D-based models of growth, i.e. the Solow critique, applicable to the models of growth without scale effects examined above.[29] Implementing a mechanism to remove scale effects, and subsequently conclude that the model performs well when compared to the empirical evidence on growth without scale effects, is not the right way to go about explaining growth without scale effects. Stated differently, some of the models seem to have taken Jones' (1995b) conclusion that "… if an endogenous growth model predicts that permanent movements in some variable X have permanent effects on growth, then […] some other variable (or variables) must also have persistent effects on growth that offset the movements of X, in a way that is determined by the endogenous growth model." (p. 502) too literally;
3. An avenue that seems more fruitful than simply incorporating some (unsubstantiated) mechanism to remove scale effects is to incorporate well documented microeconomic phenomena and regularities. This conclusion follows from the second conclusion and regards a direction that future work on explaining the absence of scale effects in growth might take. This direction simply is the development of a model of growth that is not only consistent with the absence of a scale effect in growth but also with other empirical facts preferably at the micro level. Surely, a model that partly addresses the Solow critique is both preferred over - and more realistic than - a model that lacks such a microeconomic foundation. From the models examined above, those developed by Kortum, Segerstrom, Peretto and Peretto and Smulders, seem to do best in this regard. Kortum and to a lesser extent Segerstrom, use a fishing-out effect to create a model that is able to explain not only growth without scale effects but also the time trend in patenting (the total number of

[29] As mentioned in the previous chapter Solow's (1994) critique on R&D-based growth is simply that "… growth theorists often just insert favorable assumptions in an unearned way; and then when they put in their thumb and pull out the vary plum they have inserted, there is a tendency to think that something has been proved." (p. 53). The straightforward rebuttal to this critique is of course that assumptions are simply needed in any attempt to further the understanding of economic reality through theoretical work.

> patents granted and the number of patents granted per researcher). More interesting, however, are the models of growth without scale effects developed by Peretto and Peretto and Smulders. Their incorporation of results from the industrial organization literature, e.g. the relation between firm size and R&D and the relation between market structure and innovation, seems to lead to growth without scale effects almost naturally. Hence, the third and probably most important conclusion of the discussion of models growth without scale effects presented in this section, is that incorporating the vast literature on diverse microeconomic topics such as the already mentioned relation between firm size and R&D and the relation between market structure and innovation is ultimately the key to understanding and credibly modelling growth without scale effects.[30]

This thesis takes up this advice and continues with the discussion of empirical evidence on the role the factors that are deemed to play the most important role in aggregate level growth (without scale effects), namely knowledge, R&D and spillovers, play at the micro level. That is, the next chapter looks for clues that help further the understanding of (the mechanism underlying) the observed absence of a scale effect in growth where logically one would be expected as the result of knowledge's nonrivalrousness in micro-level facts and phenomena.

[30] Thompson (2001b) argues that more attention should be paid to the microeconomic implications of models of growth as this provides the possibility to test these models. Other microeconomic issues that probably play a role in growth without scale effects at the aggregate level, are, amongst others, firm size and survival, entry, exit and growth of firms over the product life cycle, and inter- and intraindustry knowledge spillovers.

3 Knowledge, R&D and spillovers, at the firm level

Thus far, this thesis has approached scale effects in growth, resulting from the unique properties of knowledge, i.e. its nonrivalrousness, only at the macro (aggregate) level. The most important conclusion of the previous chapter, besides the absence of a scale effect in growth, however, was that the study of economic growth in general and that part of it concerned with scale effects in particular, should pay more attention to microeconomic facts and phenomena. The factors that are most important to the macro-level study of economic growth without scale effects are undoubtedly knowledge, R&D, and spillovers. Knowledge is created through R&D activities and it is knowledge's nonrivalrousness which leads to (intertemporal) knowledge spillovers and consequently to growth scale effects. It seems therefore logical to take a closer look at the role these factors play at the micro level, e.g. in firms and hence the title of this chapter, "Knowledge, R&D and spillovers, at the firm level". By investigating this role, ultimately the models of growth without scale effects discussed in the previous chapter (e.g. Jones 1995b, Smulders and van de Klundert 1995, Young 1998, Li 2000, and Peretto and Smulders 2002) can be tested (cf. Thompson 2001b). That is, the validity of the assumptions with regard to knowledge, R&D and spillovers, these models have made in order to dispose of the prediction of a scale effect in growth of the early models of R&D-based growth (e.g. Romer 1990, Grossman and Helpman 1991, and Aghion and Howitt 1992) can be assessed. This in turn is expected to shed light on important aspects of economic reality that ultimately hold the key to the understanding of - and the ability to credibly model - growth without scale effects.

In order to achieve the above, this chapter studies work done on relations between several variables by which a firm can be characterized, e.g. size, profitability, productivity, and so forth. The idea is that the influence some firm characteristics exert on one another is in a broad sense related to the role knowledge, R&D and spillovers, play in firms. This idea is rooted in the notion that at the aggregate level, (i) an economy can be characterized in much the same way as firms can, e.g. by its size, its level of productivity and so forth, and (ii) knowledge, R&D and spillovers play a role in the relation between characteristics like size and productivity. Hence, studying the relations between diverse firm characteristics is expected to provide insights on the issue of knowledge, R&D and spillovers, at the firm level. With these insights, subsequently the validity of the models of growth without scale effects can be assessed.

Sections 3.1 to 3.5 discuss several of the relations between firm characteristics that have been studied in the economics literature. Naturally, it makes no sense to examine just any relation. Instead, relations between firm characteristics, which are expected to be able to shed light on the issue of knowledge, R&D and spillovers, at the firm level, must be examined.[31] Therefore, each section starts with establishing a broad area of promising research before continuing with discussing concrete relations studied in these areas. That is, each of the sections 3.1, 3.2, 3.3, 3.4 and 3.5 examines relations between firm characteristics that can be shared under one broad research area. Section 3.6 summarizes the main results of the examination of the work done in the various research areas and provides a discussion and some conclusions.

3.1 Firm size and growth

A first logical area of research that might provide useful insights to the role knowledge, R&D and spillovers, play in firms is the relation between firm size and growth. This follows from simply applying the methodology used to investigate the existence of a scale effect in growth in Chapter 2 to the firm level.

The early models of R&D-based growth predict the existence of a scale effect in growth, which implies that large economies grow faster than small economies do. This is expressed in the way they model the knowledge production function. In these models new knowledge is created using a fraction (s) of the total labor force ($L_A = sL$) and the existing stock of knowledge (A) (see Jones 1999):

$$\dot{A} = \delta L_A A \tag{3.1}$$

where δ is a parameter determining the arrival rate of new knowledge (see also equation 2.2 in Chapter 2). Since the growth rate of output per worker equals the growth rate of the knowledge stock ($g_A = \dot{A}/A = \delta sL$) multiplied by a constant, a permanent change in research intensity (s) has a permanent effect on the growth rate. In addition, a large country is expected to grow faster than a small country because given a certain research intensity it uses a larger absolute amount of resources (labor L) in the production of new knowledge. Hence, the early models of R&D-based growth predict the existence of a scale effect in growth: the size of an

[31] Note that knowledge, R&D and spillovers are highly interrelated. As mentioned above, R&D activities lead to the creation of new knowledge, which due to its nonrivalrousness leads to spillovers (and ultimately to a scale effect in growth). The fact that the factors knowledge, R&D and spillovers are highly interrelated means that any relation between two (or more) firm characteristics that sheds light on the role any one of these factors play at the firm level, indirectly sheds light on the role played by all of them.

economy influences its rate of growth. Of course, this scale effect in growth has been refuted in much empirical work (see Chapter 2). Regarding the issue of knowledge, R&D, and spillovers, at the firm level, it seems logical and interesting to investigate whether a mechanism similar to that present in the early models of R&D-based does lead to a scale effect in growth at this level. If, for instance, growth does exhibit a scale at the firm level this can be seen as confirmation of the notion that knowledge is nonrivalrous.

A straightforward application of the knowledge production function to the firm-level would be that (firm-specific) knowledge is created using a fraction (s_i) of the firm's total number of employees ($L_{i,A} = s_i L_i$) and the firm's existing stock of knowledge (A_i):

$$\dot{A}_i = \delta_i L_{i,A} A_i \tag{3.2}$$

where δ_i is a parameter determining the arrival rate of new knowledge for firm i. The above equation depicts a situation in which the level of productivity (output per employee) is expected to increase at a faster rate in firms which perform relatively much R&D (high research intensity), and (similarly) in a large firm output per employee (productivity) is expected to grow faster than in a small firm because given a certain research intensity (s_i), the larger firm uses a larger absolute amount of people in the production of new knowledge. That is, when (firm-specific) knowledge is created according to equation (3.2), there is a scale effect in the growth of output per employee.[32] Hence, it provides a first promising area of research that might provide insight on the role knowledge, R&D and spillovers, play at the firm level, namely research that has examined the relation between firm size and growth. If this research shows that growth for larger firms (suitably measured; cf. Backus, Kehoe and Kehoe) is higher than growth for smaller firms, it can be concluded that at the firm level growth does exhibit a scale effect. If not, it can be concluded that growth scale effects are absent at the firm level as well as at the macro level.

Several studies have investigated this relation, e.g. Mansfield (1962) (Mansfield), Kumar (1985) (Kumar), Evans (1987a) (EA), Evans (1987b) (EB), Hall (1987) (Hall), Dunne, Roberts and Samuelson (1989) (DRS), and Dunne and Hughes (1994) (DH).

The findings of these studies with respect to the relation between growth and size (Growth-size) are summarized in Table 1. In addition, Table 1 provides information on the relation

[32] In the model developed by Smulders and van de Klundert (1995), the assumption that knowledge is firm specific, causes growth scale effects to be absent at the aggregate level. This assumption of knowledge being firm specific does not, however, exclude the possibility of a growth scale effect at the firm level. With the stock of firm-specific knowledge increasing over time, also the growth rate of this stock and consequently the growth rate of output per worker would increase over time (holding research inputs constant).

between growth and age (Growth-age), the data used (Sample), the factors/variables/etc. used in the case the results are tested for robustness (Robustness), and other interesting findings (besides the relation between growth and size and possibly growth and age) (Other). A more detailed discussion of the various studies can be found in Appendix B.

The studies surveyed are generally in agreement about the relation between firm size and growth: smaller firms grow faster than larger firms do. This result is robust over different periods, different industries, and different sample sources, and also robust to possible sample problems such as heteroscedasticity. In addition, several of the studies find that the growth rate varies with age, younger firms grow faster than older firms do, which is logically expected because in general older firms will also be larger firms. These findings generally concur with the predictions of the theoretical model of selection and industry evolution of Jovanovic (1982).

Regarding the issue of firm-level growth scale effects, i.e. does productivity grow faster in larger firms, the findings of the studies cannot be used to say anything useful about growth scale effects at the firm level. The measures of growth and size used in the studies, (percentage increase in) net assets, physical assets, equity assets, the number of employees and total sales, are simply not the measures that can be used to investigate a possible scale effect in growth at the firm level. For the above studies to be useful as potential evidence on firm-level growth scale effects they should have investigated relations like growth in total sales per employee as a function of total sales or growth in total sales per employee as a function of the physical assets.[33] There thus is a misleading difference between what is meant by the term growth: at the macro level growth refers to growth in a density variable, e.g. income per capita or output per worker (productivity), while at the firm (micro) level it refers to growth of some absolute amount, e.g. total number of employees.

Information regarding the role knowledge, R&D and spillovers, play in firms thus ought to come from studies that have explicitly investigated the relation between firm size and productivity growth or other measures similar to macro-level growth. Studies that can be shared under this category are Aw (2002), Illueca and Lafuente (2003), and Pagano and Schivardi (2003). Since, these studies are not suitable for orderly representation in a table like the studies discussed above, here these studies and there main results regarding the relation between firm size and productivity growth, and some other interesting findings are simply

[33] For the same reasons no inferences can be made from the relation between firm age and firm growth, despite the logical idea that older firms have performed more past research than younger firms and consequently have more (firm-specific) knowledge at their disposal.

Table 1: Main features of some work on the growth-size relation.

Study	Growth-size	Growth-age	Sample	Robustness	Other
Mansfield	Gibrat's law of proportionate effect does not hold	-	Steel, petroleum, and rubber tire industries	Different samples, e.g. exiting firms excluded, small firms excluded	Successful innovators grow, on average, twice as rapid as other firms ceteris paribus.
Kumar	Weak negative relation for different measures of size	-	Quoted U.K. firms form both manufacturing and non-manufacturing industries	Different periods (1960-1971 and 1966-1976) and different industries (manufacturing and non-manufacturing)	Weak negative relation between size and growth by acquisition
EA	Negative relation for 89% of the industries	Negative relation for 76% of the industries	100 U.S. manufacturing industries with data taken from the Small Business Data Base (SBDB)	-	The variability in growth rates decreases with size and age for 80% and 85% of the industries respectively
EB	Negative relation	Negative relation	20,000 manufacturing firms with data taken from the SBDB	Different samples, different periods and the impact of heteroscedasticity (results pass White's specification test)	The relation between growth and size is highly non-linear
Hall	Smaller firms grow faster than larger firms with the difference in annual growth rates between firms in the 25th and 75th size percentile equaling four percentage points	-	Publicly traded U.S. manufacturing firms	The relation found is not due to measurement errors (size) or sample attrition	A four million dollar increase in investment leads to a one percent increase in the annual growth rate while a two million dollar increase in R&D expenditures achieves the same effect
DRS	Relative to small firms, large plants have lower growth rates	-	200,000 plants that entered the manufacturing sector in the 1967, 1972 or 1977, Census of Manufactures (U.S.)	-	Multi-unit plants grow faster than single-unit plants of similar size and large single-unit plants have negative average growth rates
DH	Smaller firms grow faster than larger firms	For a given size, younger firms grow faster than older firms do	Quoted and unquoted U.K. companies	Results are not induced by sample problems	-

Source: Mansfield (1962), Kumar (1985), Evans (1987a and 1987b), Hall (1987), Dunne, Roberts and Samuelson (1989), and Dunne and Hughes (1994).

summarized.

Aw (2002)

Aw studies several interesting relations using firm-level data from nine Taiwanese industries for the years 1981, 1986 and 1991. The first is (1) the relation between a firm's size and its level of total factor productivity. In cross-sectional regression analysis and controlling for firm age, which always has a significant positive impact on the level of TFP, she finds that micro-sized firms (< 5 employees), are less productive than small firms (5-99 employees), which in turn are less productive than medium-sized firms (100-299 employees), and that large firms (≥ 300 employees) are most productive throughout the nine industries. The second relation is (2) the relation between TFP and firm growth. Including year dummies in her cross-sectional analysis, Aw finds that firms that are more productive grow faster, i.e. more productive firms increase their total number of employees faster. This result is found with the inclusion of initial firm size, which has a negative impact on firm growth, as an additional explanatory variable. This finding of larger firms, increasing their size at a lower speed than smaller firms do is of course consistent with the findings of Mansfield (1962), Kumar (1985), Evans (1987a and 1987b), Hall (1987), Dunne, Roberts and Samuelson (1989) and Dunne and Hughes (1994), discussed above. The third and most interesting relation she investigates is (3) the one between TFP growth and firm size. Aw again uses cross-sectional regression analysis and controls for, age, year dummies and the share of the firm's revenue that comes from subcontracting (indicating low sunk costs because firms can start up as small subcontractors). She subsequently finds some evidence that TFP growth increases with size. That is, TFP growth in micro-sized firms is lowest, while that in large firms is highest. The findings, however, are not robust over the nine different industries Aw studies and, in general, the growth rates of firms of different size do not differ statistically significant from each other. As mentioned, in the TFP growth regressions, age is also included as an explanatory variable and has a negative though not statistically significant effect on the growth rate of TFP.

Illueca and Lafuente (2003)

Illueca and Lafuente investigate the relation between the size of a firm and its productivity in a small sample of Spanish tile producers. For this purpose, they use Data Envelopment Analysis (DEA) to estimate the frontier of the production function in the different years of their data set and subsequently use the Malmquist productivity index as the measure of productivity. Illueca and Lafuente find that the smaller firms (< 2 times the average turnover;

63 firms) exhibit a productivity decline of on average 1.0% per year (relative to the productivity index), while in the largest firms (> 2 times the average turnover; 8 firms) productivity growth equals 0.2% per year (relative to the productivity index). They continue with attributing these productivity differentials to their sources, namely (i) technological change (producing more with the same input, or vice versa producing the same with less input), (ii) a catching-up effect (decreasing technical inefficiency), and/or (iii) an effect of scale (moving closer to the optimal firm size). The authors find that the largest firms benefit from both technological change and catching up, while they suffer from an effect of scale. The smallest firms (< ½ times the average turnover; 23 firms), in contrast, benefit from an effect of scale, suffer from technological change, while catching up plays a small negative role. Hence, Illueca and Lafuente conclude that productivity evolves differently depending on firm size.

Pagano and Schivardi (2003)

Pagano and Schivardi use data on the size distribution of firms in the manufacturing sector in several European countries to study the relation between productivity growth at the industry level and the average size of the firms active in the industry. With dummies controlling for unobserved sectoral differences, they find a positive relation between average firm size, measured by the total number of employees working in the average place of employment of a randomly chosen worker: the co-worker mean, and productivity growth at the industry level. That is, in industries that comprise relatively large firms productivity grows relatively fast. The extent of the effect is such that when average firm size increases with ten percent, the rate of growth of productivity increases with roughly 0.18%. The positive relation holds for several tests of robustness, e.g. the use of different measures of size and repeating the analysis for samples of different country-composition, but loses its statistical significance when country dummies are included. This last result is in line with their finding of large cross-country differences in the size distribution of firms. To further test the robustness of the relation between average size and productivity growth at the industry level Pagano and Schivardi perform tests of causality using different measures of industry-level R&D intensity and an interaction term of firm size and R&D intensity. When this interaction term of firm size multiplied by R&D intensity at the industry level is included in the previous estimation equation, size itself loses its significance and its coefficient equals approximately nil while the interaction term contributes positively and significantly to productivity growth. What is more, the results are significant even though country dummies are also included in the regression

equation. The logic behind this result is that the interaction term catches cross-country variations in firm size, which in the previous regression analysis were caught by the country dummies. In addition, the authors find that for industries with higher R&D intensity, the effect of average firm size on growth is larger. From these findings on the robustness of the relation between average firm size and industry-level productivity growth, Pagano and Schivardi conclude that it is average firm size that, quite possibly through its effect on R&D intensity, has a causal impact on growth (and not the other way around).

In general, the studies investigating the firm size productivity growth relation find that smaller firms are associated with lower productivity growth relative to larger firms. However, one needs to be careful when drawing inferences from these studies. This is due to:

1. the limited amount of studies available (especially when compared to the number of studies investigating the relation between firm size and firm growth),
2. the fact that the studies have not all examined a direct relation between firm size and firm growth,
3. the use of limited data sets which makes it ill-advised to generalize the findings of the studies.

The disappointing conclusion of this section consequently is that although the data used in the studies of the relation between firm size and firm growth discussed here does allow for the easy construction of suitable measures (cf. Backus, Kehoe and Kehoe) to test for the presence of growth scale effects at the firm level, as of yet, no conclusive evidence is available on differences in the growth rate of productivity between large firms and small firms. With useful data within arm's reach (e.g. the Small Business Data Base), investigating the relation between size and productivity growth, seems a high-yield way to gain useful insight on the issue of knowledge, R&D and spillovers, at the firm level, and ultimately perform some sort of test of the models of growth without scale effects discussed in the previous chapter.

3.2 R&D and pure knowledge spillovers

Specifying the production of knowledge at the firm level by straightforwardly adapting an already simplified version of the knowledge production function central in the early models of R&D-based growth, as is done for obtaining equation (3.2) in the previous section,

necessarily leaves out several factors that would only cloud the issue. An important complicating factor that is missing in the firm-level R&D process depicted in equation (3.2) is that it does not allow for interfirm (pure) knowledge spillovers.[34] In addition, the extent of intertemporal R&D spillovers in the knowledge production function is arbitrarily set equal to 1 (cf. Jones 1995b). A more realistic production function of (firm-specific) knowledge is:

$$\dot{A}_i = \delta_i L_{A,i} (A_i)^{\phi_i} A^{\phi_A} \tag{3.3}$$

where ϕ_i measures the extent of pure knowledge spillovers (see Appendix A) and the consequent intertemporal R&D spillovers (the degree of externalities in the R&D process across time), A denotes the total stock of public knowledge[35] (see also equation 2.4 in Chapter 2) and ϕ_A reflects the usefulness of pure knowledge spillovers (again see Appendix A) from the total stock of public knowledge. Although the above equation is still a very simplified representation of the way in which firms create knowledge, it is much more insightful than equation (3.2). Equation (3.3) shows that spillovers may (can) play an important role in the production of (firm-specific) knowledge by firms. As mentioned in the previous chapter, there is ample empirical evidence that testifies to the importance of (knowledge) spillovers at the aggregate level (e.g. Coe and Helpman 1995, Coe, Helpman and Hoffmaister 1997, Verspagen 1997, and Frantzen 2002). However, as Frantzen mentions, the strength of these spillovers is not large enough to lead to a scale effect in growth. This finding of course supports the models of growth without scale effects developed by for example Jones 1995b, Smulders and van de Klundert 1995, Young 1998, Li 2000 and Peretto and Smulders 2002, which dispose of the prediction of the early models of R&D-based growth that there is a scale effect in growth by limiting the extent of knowledge spillovers. Not mentioned in the discussion of the empirical work in the previous chapter, however, is interesting work by Porter and Stern (2000) (see also Furman, Porter and Stern 2002). This work has explicitly investigated a production function of knowledge at the country level very similar to the production function of (firm-specific) knowledge depicted in equation (3.3) and the role of the parameter(s) measuring the extent of knowledge spillovers, e.g. ϕ_i, therein.

[34] As mentioned in Chapter 2, there are two types of spillovers, pure knowledge spillovers and rent spillovers, and only the first type is directly related to knowledge's nonrivalrous nature and of interest in this thesis. Because Appendix A provides an extended discussion of the two types of spillovers, it should be viewed as important background material to the studies discussed in this section.

[35] The total stock of public knowledge could simply be defined as a fraction of firm-specific knowledge summed over all firms or as knowledge which previously was firm-specific but now has become public for instance because of the expiration of its patent protection.

Hence, although the work of Porter and Stern is not, like the other work discussed in both this chapter and this section, focussed at the firm level and in fact it would not be out of place in the discussion of empirical evidence on growth scale effects in the previous chapter, this section starts with a discussion of this work on R&D, spillovers and the production of knowledge, in Paragraph 3.2.1. Having discussed the country-level work of Porter and Stern, the section continues with examining firm-level evidence on pure knowledge spillovers between firms (interfirm) and within firms (intrafirm) respectively in the next two paragraphs (3.2.2. and 3.2.3). Some remarks to conclude the section are presented in Paragraph 3.2.4.

3.2.1 Country-level knowledge production

Starting point for Porter and Stern's work on the production of knowledge by countries is the following production function of new-to-the-country (country-specific) knowledge (A_j):

$$\dot{A}_j = \delta(L_{A,j})^{\lambda}(A_j)^{\phi}(A_{-j})^{\psi} \tag{3.4}$$

where λ is a parameter, A_{-j} denotes the world stock of public knowledge minus the country's own knowledge stock, ϕ is a parameter measuring the extent of knowledge spillovers and the consequent intertemporal R&D spillovers (the degree of externalities in the R&D process across time) with respect to the country's own knowledge stock (A_j), and ψ is a parameter measuring the extent of the spillovers from the world knowledge stock.[36] The knowledge production function depicted above is obviously closely related to the (firm-specific) knowledge production function depicted in equation (3.3) and the generalized knowledge production function of Jones, Kortum and Segerstrom, depicted in equation (2.4) in Chapter 2. The only difference is the adding of the parameter λ. Adding the parameter λ to the models of Jones, Kortum and Segerstrom, does not qualitatively alter their most important features; growth is still semi-endogenous and still does not exhibit a scale effect. The only difference is quantitatively in that adding this parameter changes the growth rate of productivity along the balanced growth path from $g_A = n/(1-\phi)$ to $g_A = \lambda n/(1-\phi)$. In equation (3.4) the parameter $\lambda < 1$ can best be seen as reflecting a possible (negative) duplication externality: some of the new knowledge a person, e.g. a researcher, creates is not new to the world as a whole (knowledge output is concave in R&D effort). Jones (1998a) refers to the effect of $\lambda < 1$ as the 'stepping-

[36] Of course, for very large countries, that is countries which knowledge stock is not negligible relative to the world knowledge stock and which contribution to the world knowledge stock is not negligible relative to the total worldwide contribution, ψ, similar to ϕ, denotes the extent of an intertemporal knowledge spillover.

on-toes' effect. Of course in real life duplication seems important simply because firms compete with each other to be the first to develop certain new and/or improved products, e.g. a DVD recorder, LCD screens, etc.

Porter and Stern use the knowledge production function of equation (3.4), as the basis for their estimation equation. The knowledge data they use is data on international patents, i.e. patents granted by the U.S. patent office to foreign establishments, and the data on R&D resources is simply the number of full-time-equivalent scientists and engineers. They subsequently model a country's patent output ($\dot{A}_j$) (three periods in the future) as a function of the (current) number of scientists and engineers ($L_{A,j}$), the (current) country's Own patent stock (A_j) and/or the (current) Rest of world patent stock (the total world patent stock excluding the country's own patent stock; $A - A_j = A_{-j}$) (A_{-j}). In total Porter and Stern estimate three different basic knowledge production functions, comprising one or more of these variables as explanatory variables, the main results of which are depicted in Table 2 below.

Table 2: Country-level knowledge production.[37]

Dependent variable: ln(patents)	1			2		3	
ln(S&E)	0.48	0.39	0.85	0.81	0.30	0.36	0.20
ln(Own patent stock)	1.19	0.84	0.02•	1.03	0.90	0.85	0.94
ln(Rest of region patent stock)	…	…	…	…	…	-0.01	…
ln(Rest of shared language stock)	…	…	…	…	…	…	-0.02
ln(Rest of world patent stock)	…	…	…	-1.05	-1.01	-0.58	-0.28
Baseline year dummy	…	1.12	…	…	0.17•	0.56•	0.94
Year trend	…	-0.08	…	…	0.03	-0.01•	-0.05
ln(GDP in 1967)	…	-0.24	…	…	-0.21	-0.26	-0.15
Year dummies exist?	Yes	No	No	No	No	No	No
Country dummies exist?	Yes	No	Yes	Yes	No	No	No
R^2	0.99	0.98	0.98	0.99	0.98	0.98	0.98

Remarks: (•) denotes insignificance.

They present remarkable results, not the least of which is that all the different models they use in each of the three basic specifications are able to explain nearly all of the variance in patent output. Individually the three basic knowledge production functions reveal some interesting things as well.

1. The first basic knowledge production function is a simple model with only the number of scientists and engineers (S&E) and the Own patent stock (A_j) included as explanatory variable (see Romer 1990, Grossman and Helpman 1991, Aghion and Howitt 1992, Jones 1995b, Kortum 1997, and Segerstrom 1998). The most important results of this model,

[37] Source: Porter and Stern (2000).

besides the fact that it is able to explain a large portion of the observed variance in patent output, is that it shows that (i) there is a large duplication externality, i.e. the returns to research effort decrease with increasing R&D effort, as evidenced by the value for λ of 0.48, and that (ii) the extent of the intertemporal R&D (knowledge) spillover is large as well ($\phi = 1.19$) and the null-hypothesis that $\phi < 1$ can be rejected. A first conclusion thus is that the production of knowledge (as measured by patents) is better described by the knowledge production function of the early models of R&D-based growth (Romer, Grossman and Helpman, and Aghion and Howitt) than by the one-R&D-sector models of growth without scale effects (Jones, Kortum, and Segerstrom). Porter and Stern also estimate the first basic estimation equation without the full set of year- and country-dummies, and only a Baseline-year dummy, a Year trend and a country-specific variable (GDP in 1967) included as control variables. They find that the sensitivity of patent output to R&D resources decreases to 0.39, which does not differ significantly from the earlier value of 0.48. The sensitivity of patent output to the patent stock decreases significantly as well to 0.84. Porter and Stern attribute this decline in ϕ to the inclusion of the country-specific effects (GDP in 1967) and this is confirmed when next to the number of scientists and engineers and the own patent stock, only country dummies (i.e. no Baseline-year dummy, no Year trend and no country-specific variable) are used as independent variables in the regression equation. In this regression, the value of ϕ declines to 0.02, which does not differ significantly from zero, while λ rises to 0.85. Finally, Porter and Stern find a downward trend in patenting productivity of 8% per year. There thus seem to be two time effects working counter to each other: a positive one resulting from the patent stock increasing over time and a negative one related simply to the passage of time.

2. In the second basic knowledge production function, the authors estimate not only the number of scientists and engineers (S&E) and the Own patent stock (A_j) are included as explanatory variables but also the Rest of world patent stock (A_{-j}). Controlling for country-specific effects through the inclusion of country-dummies they subsequently find values of 0.81 and 1.03 for λ and ϕ respectively while the rest of the world patent stock has a strong negative net effect ($\psi = -1.05$). Similar results are found when this second basic specification of knowledge production is estimated with a Baseline year dummy, a Year trend and GDP in 1967 as control variables rather than a full set of year- and country-dummies. However, the duplication externality becomes more severe ($\lambda = 0.30$) and the passage of time now has a positive effect while in the first basic specification

(without the rest of world knowledge stock as explanatory variable) it had a negative effect. A possible explanation for these results is that the world knowledge stock has two effects working counter to each other (see also the results obtained with the first basic knowledge production function discussed above). There is a positive effect of the world knowledge stock which helps to increase patent output through spillovers (the standing-on-shoulders effect; see Jones 1998a) and there is a negative effect, which Porter and Stern refer to as the 'raising-the-bar' effect (cf. Jones' 1998a fishing-out effect), which means that increases in the number of ideas discovered and patented by the other countries combined diminishes a country's ability to discover and patent knowledge that is new to the world (p. 22). The negative effect obviously has the upper hand.

3. For the third basic knowledge production function Porter and Stern divide the Rest of world patent stock (A_{-j}) in the patent stock of either the rest of the region to which a country belongs or the rest of the shared language group to which a country belongs, and the remaining rest of the world patent stock.[38] Hence, in the third basic specification of the knowledge production function, patent output is modelled with the number of scientists and engineers (S&E), the Own patent stock, the remaining Rest of world patent stock and either the Rest of region patent stock or the Rest of shared language patent stock as explanatory variables. With a Baseline year dummy, a Year trend and GDP in 1967 included as control variables, they find that neither the regional nor the shared-language patent stock have a strong impact on patent output. The impact of both the Own patent stock (ϕ) and the Rest of world patent stock remains significant however.

The conclusions Porter and Stern draw from the findings discussed above are that:

- The domestic knowledge (patent) stock contributes positively and significantly to the production of new-to-the-world knowledge. In fact, they conclude that the extent of the standing-on-shoulders effect is such that the productivity of a country's R&D efforts increases proportionally to the stock of accumulated knowledge ($\phi \geq 1$);
- Spillovers with international origin, as opposed to those with domestic origin, have a negative impact on a country's production of knowledge that is new to the world. That is, internationally, the raising-the-bar effect is stronger than the standing-on-shoulders effect so that the net contribution of the world knowledge (patent) stock to national knowledge

[38] The regions are: Australia & New Zealand, Austria, France, Germany, Italy, the Netherlands, Spain & Switzerland, Denmark, Finland, Norway & Sweden, Canada & the U.S., and Japan and the U.K. are ungrouped. The language groups are: Australia, Canada, New Zealand, the U.K. & the U.S., Austria, Germany & Switzerland, and all other countries are ungrouped.

production is significantly negative;

- There is a strong stepping-on-toes effect: patent output increases less than proportionally with increases in the people (scientists and engineers) engaged in R&D.

The conclusion of Porter and Stern (2000) regarding the extent of intertemporal R&D spillovers is strongly at odds with the conclusion Jones (1995a) reaches regarding this spillover on the basis of data on observed steady (TFP) growth and the steady increase of growth promoting factors like R&D employment and the question is how these two seemingly contradictory findings can be reconciled. For this purpose, Porter and Stern investigate the relation between the knowledge stock as measured by international patents ($A = A_j + A_{-j}$) and the level of GDP and TFP. They find that a doubling of the patent stock leads to an increase of only 22% in the level of GDP and that, similarly, the level of TFP increases only 5% in response to a doubling of the patent stock (A), which leads them to conclude that the growth effect of knowledge production are modest, limited by, amongst others, the stepping on toes effect (i.e. $\lambda < 1$; concavity in research effort). Stated differently, there seems to be a discrepancy between countries' ability to produce new knowledge (as evidenced by patents) and their ability to put this new knowledge to productive use.

Further issues arise however.

1. First of all regarding the relation between patents and growth of (total factor) productivity or income. Surprisingly, this relation has not been widely studied, as opposed to the relation between R&D (input) and patents (output), and R&D and growth of (total factor) productivity or income respectively (see for instance Pakes and Griliches 1984, Griliches and Lichtenberg 1984 and other studies in Griliches 1984). The few other studies that have investigated the patent-growth relation, e.g. Lach (1995), Verspagen (1996) and Bloom and van Reenen (2002) do, however, present findings similar to that of Porter and Stern (2000). Lach (1995) uses industry-level data and finds an elasticity of industry TFP with respect to the (industry) knowledge stock (constructed using patents actually used in the respective industries) of 0.30.[39] Verspagen (1996) studies the effect of growth of the knowledge stock on labor productivity at the country level and finds various interesting results. In his estimates for Germany, France, and the U.K., he finds that the elasticity of country labor productivity with respect to the country knowledge stock (again constructed

[39] Lach also investigates the relation between a knowledge stock constructed from past R&D expenditures and TFP, and finds an elasticity of TFP with respect to this knowledge stock of ± 0.085 when only company-financed R&D is included, and an elasticity of ± 0.049 when all industry-level R&D is included in the construction of the knowledge stock. So R&D and TFP are even less strongly related than are patents and TFP.

using patents) varies between 0.026 for the U.K. (not significantly different from zero), 0.150 for Germany, and 0.253 for France. Using a different patent lifetime, 10 years as opposed to 15 years, in constructing the knowledge (patent) stock yields much the same results though the elasticities are generally slightly lower: 0.037 for the U.K. (not significantly different from zero), 0.126 for Germany, and 0.232 for France. In a cross-country regression with other countries included, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the U.K. and the U.S., Verspagen finds similar results: when the long patent lifetime is used the estimated elasticity of (country) labor productivity to the (country) knowledge stock equals 0.118, and when the short patent lifetime is used, the estimated elasticity equals 0.088.[40] Finally, Bloom and van Reenen (2002) study the effect of growth of the knowledge stock on TFP at the firm level and also find various interesting results. In a sample consisting of U.K. firms that have actually taken out patents, they find an elasticity of firm TFP with respect to the firm knowledge (patent) stock of about 0.03. When subsequently a knowledge stock based on patent citations, a proxy for the quality of patents (cf. Griliches 1990) is included, the patent stock loses its significance while the elasticity of TFP with respect to this citation stock is 0.038, slightly higher than that of the patent stock. Hence, the few other studies that have investigated the patent-growth relation confirm the findings of Porter and Stern of a limited relationship between knowledge production and productivity (growth). The actual magnitude of the effect knowledge production has on productivity (growth) unfortunately is unclear. The different studies report elasticities of (total factor) (labor) productivity with respect to patents varying from 3% to 30%. Hence, a direction for future research would be to further investigate the relation between growth of the knowledge stock, as measured by, for instance, a patent stock, and growth (of per-capita income, TFP and/or labor productivity).

2. The second issue that arises from Porter and Stern's original finding of the limited effect the production of knowledge has on TFP (or per-capita income), regards the reason underlying this apparent discrepancy between countries' (or firms') ability to produce new

[40] Similar to Lach, Verspagen has investigated (some of) the relations using an R&D-based knowledge stock instead of a patent-based knowledge stock. He also finds that R&D and labor productivity are even less strongly related than are patents and labor productivity: the regression yields estimates of the elasticity of country labor productivity with respect to the country R&D-stock of 0.054 for the U.K. (not significantly different from zero), 0.157 for Germany, and 0.098 for France when a perpetual inventory measure is used to construct the R&D stock (comparable to a long patent lifetime), and an elasticity of 0.013 for the U.K. (not significantly different from zero), 0.096 for Germany, and 0.077 for France when a moving average with a bell-shaped weighting structure is used to construct the knowledge (R&D stock) (comparable to a short patent lifetime).

knowledge and their ability to translate this knowledge in productivity (growth). TFP is a measure of the effectiveness knowledge has in. A possible answer is presented in the findings of Lach (1995), Verspagen (1996), and Bloom and van Reenen (2002), and the discussion of the usefulness of patents in the study of economic growth in Griliches (1990). The idea is that it is difficult to measure the knowledge stock (see also Chapter 5): R&D expenditures are really only an input to the process of knowledge production, and patents, while having the advantage of being an obvious output measure, have the disadvantage of not accurately reflecting the value of the underlying knowledge. This idea is confirmed in the work discussed above: the relation between knowledge growth and growth (of per-capita income, TFP or labor productivity) is stronger when a patent-based measure of the knowledge stock is used rather than an R&D-based measure (Lach and Verspagen). Similarly, a knowledge stock based on patent citations, a logical measure of the value (or quality) of the knowledge underlying a patent (cf. Griliches 1990), has a closer relation to productivity than a knowledge stock based on (just) patents (Bloom and van Reenen). Besides the concavity in research effort (the stepping on toes effect), there thus may be another reason underlying the weak relation between patents and productivity, namely the inability to measure (the value of) knowledge accurately. That is, the contradictory findings of Porter and Stern (2000) and Jones (1995a) regarding the extent of intertemporal R&D spillovers may be due to measurement difficulties.[41]

Chapter 5 on measurement issues in the study of R&D-based growth, presents a broader discussion of the measurement of knowledge and discusses further possible explanations for the weak relation between patents (or more broadly measures of knowledge) and productivity (growth). The conclusion drawn from the discussion here and ending this section is that future work should not only further investigate the relation between growth of the knowledge stock and growth (of per-capita income, TFP and/or labor productivity), but should also experiment with different measures of the knowledge stock, e.g. citation-based. Undoubtedly, repeating the work of Porter and Stern with citation-based measures, will render different results including a stronger relation between knowledge (growth) and productivity (growth).

[41] Griliches (1990) discusses the use of patents as economic indicators, for instance as measures of knowledge and points out that patents are not a perfect measure, for instance because the value of the knowledge underlying a patent varies per patent, and citations go only so far in laying bare this value. It is important to keep in mind this weakness of patents as economic indicators for in several instances the studies discussed in this thesis have used patents as their measure of knowledge, e.g. Kortum, Segerstrom and of course Porter and Stern. See also the studies in the next paragraph and the next section.

3.2.2 Interfirm pure knowledge spillovers

Several studies have investigated interfirm pure knowledge spillovers, e.g. Jaffe (1986), Jaffe (1988), Megna and Mueller (1991), and Acs, Audretsch and Feldman (1994). Unfortunately, the variety of these studies does not allow orderly representation of the most important findings of these studies in a table like the studies of the firm size firm growth relation discussed in the previous section. Hence, here a short summary of the studies and their most important results is presented while Appendix C.1, by virtue of the interesting findings of these studies, provides a considerably more detailed discussion of them.[42]

Jaffe (1986)

Jaffe investigates the effect technological opportunity and R&D performed by other firms (R&D spillovers) has on a firm's R&D productivity. He constructs a measure of technological distance (see also Griliches 1979) and finds that interfirm (pure) knowledge spillovers are significant: the productivity of a firm's R&D process is increased by the R&D efforts of its 'technological neighbors' (firms that are close by in 'technology space'). R&D conducted by competitors, however, has both a negative direct and a positive indirect effect on a firm's profits and market value. The direct effect being negative means that the more R&D competitors perform, the lower both a firm's profits and its market value are. The indirect effect is that firms that perform relatively much R&D themselves, can benefit from large spillovers[43] and consequently increase both their profits and their market value. For firms that perform relatively little R&D, the combined effect is unambiguously negative. Regarding the effect technological opportunity has on a firm's R&D productivity, no definite conclusions can be drawn.

Jaffe (1988)

Using much the same methodology as Jaffe (1986), Jaffe (1988) investigates the effect (i) technological opportunity and R&D (knowledge) spillovers (supply factors) and (ii) market

[42] Appendix A of course provides important background information on the difference between (pure) knowledge spillovers, which are the spillovers that are of interest here and in this thesis, and rent spillovers, which are much less associated with knowledge's nonrivalrousness, and consequently of considerably less interest.

[43] This is consistent with the idea of Cohen and Levinthal (1989) that firms perform R&D to increase their ability to use outside knowledge, i.e. to increase their learning or absorptive capacity. More R&D effort by other firms increases the spillover pool, and a firm can make more use of this pool by increasing its learning and absorbing abilities through R&D investment.

conditions (demand factors), have on R&D intensity and productivity growth. He finds results that are consistent with the results of his earlier study (Jaffe 1986). That is, both a firm's R&D intensity and its rate of TFP growth increase in response to (see footnote 43; cf. Cohen and Levinthal 1989) - respectively increase with - the R&D efforts of other firms. In addition, Jaffe (1988) finds that in the R&D intensity equation, more than half of the spillovers emanate from outside the firm's technological cluster, while in the TFP growth equation the impact from spillovers emanating from outside the firm's technological cluster on TFP growth is negligible; TFP growth does not increase with an increase in out-of-cluster R&D. Furthermore, both technological position (technology push) and industry effects, e.g. the position of the firm in 'market space' (market pull), are found to be statistically significant in explaining cross-firm variance in R&D intensity, while they are only significant in explaining TFP growth when the potential spillover pool is not included in the estimating equation.

Megna and Mueller (1991)

Megna and Mueller try to explain cross-firm differences in profit rates from differences in firms' intangible capital endowments, i.e. from differences in their R&D capital and advertising capital.[44] They find that other firms' advertising capital generally leads to lower sales and profits, which means that the competitive effect of rivals' advertising dominates the positive externality of rivals' advertising. Industry R&D in turn can have both significant negative and significant positive effects on a firm's sales and profits, and this varies over the four industries examined.

Acs, Audretsch and Feldman (1994)

Finally, Acs, Audretsch and Feldman investigate how receiving (pure) knowledge spillovers from R&D conducted elsewhere varies with firm size. The authors find that for large firms the R&D performed by other firms in the industry is the more important source while for small firms university research is the more important source of R&D spillovers.

[44] Advertising and R&D are similar type of investments: firms invest in these and create a temporary monopoly that allows them to make a profit and earn back their investment. What is more, a firm which through advertising has increased demand for its products so that it commands a price premium, has increased its productivity in much the same way as an investment in R&D might have done. The difference between R&D and advertising is that it is improbable that past advertising expenditures (advertising capital) increase the effectiveness of future advertising expenditures. However, advertising does have an externality in that advertising for one product may increase demand for products that fulfil the same role (use toothpaste to prevent caries), or for complementary products (coffee and sugar). Advertising thus has only static positive externalities while R&D has both static and dynamic ones. The issue of dynamic vs. static externalities is raised again in the final section of this chapter.

3.2.3 Intrafirm pure knowledge spillovers

Several studies have investigated intrafirm knowledge spillovers, e.g. Adams and Jaffe (1996), Henderson and Cockburn (1996), and Klette (1996). Again, the variety of these studies does not allow for orderly representation of their findings in a table. Below the studies will be summarized and their most important results discussed. Because, similar to the studies of interfirm spillovers, their findings are quite interesting, Appendix C.2 provides a considerably more detailed discussion of them.

Adams and Jaffe (1996)

Adams and Jaffe study transfers of knowledge across different plants within a single firm and spillovers of knowledge across firms. They find a lot of very interesting results. The first is that there is a life-cycle effect, with measured productivity being lower both in new plants and in (older) plants that are cutting back[45]. Relative plant age also matters, each year of age relative to the firm's average plant age is associated with approximately 1% higher productivity. The most interesting findings, however, are the findings on the way in which R&D or more broadly, knowledge, contributes to productivity (both at the plant and at the firm level). Adams and Jaffe find that the total number of plants that constitute the parent firm has a highly significant negative impact on the level of productivity of its plants. This finding holds even when firm size is controlled for. In fact, firm size is found to have a significant negative impact on plant level productivity but loses its significance when the number of plants is included in the regression. Furthermore, they find that (pure) R&D spillovers both within and between firms are very important, and that geographical distance does not impede them. The impact of spillovers is, however, diluted by the number of receivers, i.e. by the number of plants a firm comprises. Overall, their conclusion thus is that it is not the absolute amount of R&D that matters for productivity but rather its relative amount, i.e. its intensity. This of course is in conflict with the predictions of the early models of R&D-based growth, and dilution of knowledge (R&D) seems to be an important countervailing force to the increasing returns associated with the nonrivalrous nature of knowledge. Adams and Jaffe propose that their findings can be explained by technological distance, which to some extent is always present and makes it impossible to use knowledge created by other plants and/or firms without some adaptation being required.

[45] Life cycle-effects are defined in this thesis as (relative) age of a firm or plant having an effect on its economic characteristics, e.g. size, size growth, productivity growth, and so forth.

Henderson and Cockburn (1996)

Henderson and Cockburn examine the relation between a firm's size and its research (not development) productivity. They find ample evidence corroborating their hypotheses that (i) individual research programs are not subject to returns to scale, that (ii) there are returns to scale in research at the firm level, that (iii) there are significant returns to scope in research, and that (iv) research productivity is positively associated with (pure) knowledge spillovers between firms. In addition, Henderson and Cockburn find evidence that supports the idea that a change in the way research is conducted, can change the importance of scale, scope and spillovers, in the research process.

Klette (1996)

To conclude, Klette investigates the relation between R&D and his measure of firm/plant (productivity) performance. He finds that performance is a function of previous year performance and the total amount of R&D expenditures of the firm. Most of the cross-plant variance in performance could be explained by previous year performance, indicating that performance differences may be persistent over time. The type of plant was also found significant in explaining differences in plant performance. In addition, Klette investigates the effect of spillovers across plants active in the same line of business and belonging to the same interlocking group of firms and finds that the effect of these spillovers is strong. Finally, he investigates the effect of spillovers across lines of business both within the firm and within the interlocking group. There is evidence of a positive relation between plant performance and R&D effort in other lines of business within the firm but the effect is insignificant. The effect of R&D effort in other lines of business within the group, in contrast, is both positive and significant.

3.2.4 Concluding remarks

The studies of Porter and Stern (2000), Jaffe (1986), Jaffe (1988), Megna and Mueller (1991), Acs, Audretsch and Feldman (1994), Adams and Jaffe (1996), Henderson and Cockburn (1996) and Klette (1996), have provided an enormous amount of information on pure knowledge spillovers.

They unanimously conclude that these spillovers are existent and significant. Although this indicates that knowledge is at least partly nonrivalrous, the real question is of course whether the extent of these spillovers is such that they may lead to a scale effect either at the firm

level, or at the aggregate level, i.e. in growth. On this point, the country-level evidence is most clear. Porter and Stern conclude that there is an intertemporal (growth) scale effect in the production of new knowledge at the country-level. That the effect of this knowledge growth scale effect on economic growth is limited, as evidenced by for instance Jones (1995a and 1995b), follows from the fact that the production of new knowledge as measured by international patents has almost no relation to growth of either per-capita GDP or TFP (see also Lach 1995, Verspagen 1996 and Bloom and van Reenen 2002). Unfortunately, the information the firm-level studies of R&D and knowledge spillovers provide on the question whether the extent of these spillovers is such that they may lead to a growth scale effect at either the firm level, or at the aggregate level, is less clear-cut. Some of these studies find that there are benefits of being large and being active in several areas to be reaped. Other studies on the other hand find that no such benefits exist. The most interesting study is the one by Adams and Jaffe. These authors find that knowledge is diluted when it is spread over more subjects, even when these subjects are (i) 'close', i.e. technologically similar to the subject that originally created the knowledge, (ii) part of the same firm, and/or (iii) geographically near to the subject that originally created the knowledge. The dilution of knowledge is the natural way to ultimately explain growth without scale effects, for it signifies that it is not the absolute amount of resources spent on the creation of new knowledge (R&D), but rather the amount spent per economic subject (e.g. a firm, a plant, etc.). Peretto and Smulders indeed have used the concept of dilution and combined it with technological distance to create a two-R&D-sector model of growth without scale effects (see Chapter 2). Hence, thus far the Peretto-Smulders model performs best in a hypothetical microeconomic test of models of R&D-based growth without scale effects.

Based on the studies discussed above and their findings, several issues provide interesting subjects for further study. More information on some of these issues may in turn shed light on the role knowledge, R&D and spillovers, play at the firm level.
The study of Porter and Stern, but also (and perhaps even more so) the studies by Lach, Verspagen, and Bloom and van Reenen, call, as mentioned above, for more work on the relation between growth of the knowledge stock, constructed using different measures, and growth (of per-capita income, TFP and labor productivity). More concrete it would be interesting to (i) extend the firm-level study of Bloom and van Reenen (2002) with spillovers (and spillover pools) and other features of the studies by Jaffe (1986 and 1988), to estimate a firm-level production function along the lines of equation (3.3) using patent citations as the

relevant measure of knowledge, and to (ii) further investigate the reason(s) underlying the contradictory findings of Porter and Stern (2000) and Jones (1995a and 1995b) regarding the extent of intertemporal R&D spillovers (see also Chapter 5).

Jaffe's (1986) study brings forth the question whether there are clusters of technologically related firms in which knowledge is not diluted, and whether growth within these clusters does exhibit a scale effect.

The somewhat scattered evidence on the effects of R&D performed within the same industry (negative or positive) found by Megna and Mueller raises the question whether there is a hump-shaped relation between proximity and the contribution of R&D spillovers to the recipient firm. R&D performed by technologically very similar firms may have a negative effect through its effect on competition. R&D performed in firms less technologically related may, however, have a positive effect through relatively useful spillovers and limited competitive effects, while R&D efforts by technologically unrelated firms simply provide no useful spillovers.

The study by Adams and Jaffe provides, unsurprisingly, the most directions for future research.

1. The first is the relation between relative age and productivity they find. This finding is consistent with the combination of the results of Evans and Dunne and Hughes who find a positive relation between age and size, and Aw who finds a positive relation between size and the level of productivity.[46] It would now be interesting to know the effect of age and size on other variables figuring prominently in Adams and Jaffe's study. It seems logical that larger firms in general also consist of a larger number of plants relative to smaller firms (see footnote 46), which in turn would imply that older firms comprise of more plants. More broadly, age may have a large influence on a range of firm characteristics including its level of productivity, its size, and the number of plants it comprises.
2. Another interesting direction for future research is to investigate whether the number of plants has another effect besides lowering productivity. Firms are expected to maximize profits, and it would therefore be interesting to know how spreading production over a larger number of plants, is consistent with this goal. Investigating the relation between a firm's number of plants and its performance, e.g. profitability seems logical. For as far as

[46] Adams and Jaffe conversely find that plant-level productivity and firm size are negatively related, while, as stated, they also find a positive relation between relative plant age and productivity at the plant level. The reason behind these seemingly contradictory findings seems to be that in Adams and Jaffe's study firm size has a strong positive correlation to the number of plants. In addition, the effects Adams and Jaffe find is for relative plant age, which differs from absolute age.

the number of plants is in part the result of past firm behavior, it may reflect for instance past diversification moves (e.g. mergers and acquisitions) a firm has taken to maximize profit or to survive.

3. The most interesting direction for future research, however, is to determine at what level dilution of knowledge starts. Does it start when knowledge is transferred between employees of a firm? Or when it is transferred from one department of the firm to another? A concrete question would be whether knowledge is diluted in a firm that operates only one plant, is active in a single industry, produces a single product, etc., for a considerable time, i.e. does such a firm exhibit a scale effect in growth?

The study by Henderson and Cockburn has shown that the importance of scale, scope and knowledge spillovers, changes over time. It would subsequently be interesting to know what makes for an increase of the contribution of scale, scope and knowledge spillovers, to productivity and productivity growth both in a single firm and in the economy at large. Beforehand it seems logical that the importance is enhanced by increasingly codifying (previously tacit) knowledge (cf. Nelson and Winter 1982). That is, the number of plants (receivers) is expected to matter less the more codified knowledge is.

Klette has found some evidence of differences in plant performance related to the type of plant. It therefore seems interesting to incorporate data on plant type in statistical analysis of knowledge dilution like that done by Adams and Jaffe.

To conclude this section, a final and most important hint for future research is to apply the methodology used by Porter and Stern to investigate the adapted knowledge production function at other levels of investigation as well. Testing some sort of knowledge production function like the one depicted in equation (3.3), to gather information on the role R&D, knowledge and spillovers, play in firms of course was the basic idea underlying this section, and it indeed seems to provide a fruitful avenue of research. Testing knowledge production functions can, however, be done at several levels of economic analysis, at the firm level (e.g. Bloom and van Reenen and Jaffe), but also at the regional level, or at level of the world as a whole, and different measures of knowledge can be used, e.g. patents, citations and so forth. The results for each of these levels and different measures of knowledge, can subsequently be compared with the knowledge production functions of the macro growth models discussed in the previous chapter, e.g. Romer-Grossman-Helpman-Aghion-Howitt vs. Jones-Kortum-Segerstrom.

3.3 Knowledge, R&D, spillovers, increasing returns, and the firm

The first two sections have examined work done in promising areas of research that have been established by applying the methodology used in Chapter 2 to investigate scale effects in growth, to the firm level. These sections consequently have discussed the most direct information available on the role knowledge, R&D and spillovers, play in firms. However, there is probably also a lot of 'circumstantial evidence' that could shed light on this issue. Therefore, in this section an area of research expected to provide indirect information, on the role played by knowledge, R&D and spillovers, at the firm level, is established and subsequently work done in this area will be discussed.

As noted in the previous chapter, the nonrivalrous nature of knowledge gives rise to increasing returns. The most direct manifestation of these increasing returns associated with knowledge (or equivalently of the nonrivalrous nature of knowledge) at the firm level would be that knowledge gives rise to the well-known economies of scale and scope[47]. Since, in general, firms' ultimate goal is to maximize profits, it seems logical that they try to exploit any economies of scale and scope associated with its resources, including its (firm-specific) knowledge and other similar resources such as 'organizational capital' or 'unique core competencies'. Exploiting possible economies of scale and scope resulting from knowledge's nonrivalrous nature can subsequently be done best by using knowledge in the widest possible way; in the production of any given number of products of a certain type and/or in the production of any given number of other types of products.[48] Firms can - and have attempted

[47] Economies of scale can be defined as the ability of a firm to achieve lower average and marginal costs by applying its resources, e.g. capital, human capital and (firm-specific) knowledge, to a larger number of products (increasing returns to scale). Similarly, economies of scope can be defined as the ability of a firm to achieve lower average and marginal costs by applying its resources, e.g. capital, human capital and (firm-specific) knowledge, them to a wider range of products. Mathematically economies of scale can be defined as: $TC(\mu Y)/\mu Y < TC(Y)/Y$ where TC is the total cost of production, Y is the number of products produced, and $\mu > 1$. Similarly, economies of scope can be defined mathematically as: $TC(Y_1,Y_2) < TC(Y_1,0) + TC(0,Y_2)$ where TC again denotes the total cost of production, Y_1 denotes the total production of products of type 1, and Y_2 denotes the total production of products of type 2 (cf. Panzar and Willig 1981).
From the discussion of studies like Adams and Jaffe and Henderson and Cockburn in the previous section it is clear that the evidence on the existence and the extent of economies of scale and/or scope related to knowledge and its nonrivalrous nature is mixed at best.

[48] Note however, that the most profitable width of the use of knowledge is finite and limited by factors not related to knowledge and its properties (see for example Williamson 1967). For instance, as Teece notes, a firm does not have to exploit the economies of scale and scope that (might) result from knowledge (or other firm resources), itself. If the firm is able to negotiate a contract with another firm over the use of this knowledge, this might be a better, more profitable, way to reap the benefits of economies of scope and/or scale. The firm will need to pursue available opportunities to realize economies of scope and/or scale itself only if the excess resources leading to these economies are subjected to high transaction costs (indivisibilities) and subsequent market failures (see also Arrow 1962).

to - make wide use of their knowledge in two basic, though highly related ways, namely diversification (cf. Penrose 1995, and Teece 1980 and 1982) and (of course) firm size. Taking the argument one step further, it can be argued that economies of scope and scale associated with the nonrivalrous nature of knowledge partly determine the characteristics of a firm, namely the extent to which it is diversified and its size. Stated differently, an incentive effect where R&D (expenditure/effort/etc.) leads firm to diversify or to increase their size is expected. An area of research that seems promising in its ability able to shed light on the issue of knowledge, R&D and spillovers, at the firm level, would subsequently comprise work that has related knowledge, R&D and spillovers, to diversification by firms and to firm size.[49] More concrete, it would be interesting to know how knowledge, R&D and spillovers, influence diversification by firms and firm size. Unfortunately, work done in this broad area of research is limited. Nonetheless, there are some studies that have studied interesting relations, namely the relation between R&D and diversification and the relation between R&D and firm size.

This section continues with discussing the work done on the relations in Paragraph 3.3.1 (R&D-diversification) and Paragraph 3.3.2 (R&D-size) respectively. This is done by using the various studies done in this broad area of research to investigate a concrete hypothesis regarding (1) R&D and diversification and (2) R&D and firm size respectively and their relationship. At the end of the section, some concluding remarks are presented (Paragraph 3.3.3).

3.3.1 R&D and diversification

Several studies have investigated the relations between R&D and diversification, e.g. Gort (1962), Scherer (1965), Grabowski (1968), Teece (1980), Link and Long (1981), Lemelin (1982), Scherer (1983), MacDonald (1985), and Montgomery and Hariharan (1991). The results of these studies regarding the relations between R&D and diversification and other possibly interesting findings are summarized below.[50]

Before discussing these studies and their main findings, first a concrete hypothesis with respect to this relation is formulated. From the discussion above, it is clear that knowledge,

[49] It is likely that larger firms are also more diversified firms and that more diversified firms are also larger firms.

[50] Just like some studies discussed in the previous sections, these studies do not lend themselves for representation in a table. Since they are, on average, easy reading and clear on the results that are of interest here, they do not require a more detailed discussion in an appendix however.

which is created through R&D, is expected to provide an incentive for firms to diversify; knowledge's nonrivalrousness, limited as it may be, gives rise to economies of scale and scope and diversification can be used to exploit these economies (cf. Penrose and Teece). Concrete the hypothesis regarding the relation between R&D and diversification is thus that R&D leads to diversification (because diversification is a way to exploit the economies of scale and scope associated with knowledge's nonrivalrousness).
Knowing whether this hypothesis is correct or incorrect will shed light on one of the basic questions of this section (and paragraph), namely how knowledge, R&D and spillovers, affect diversification by firms. Obviously confirmation of the hypothesis, i.e. finding that R&D does lead to diversification, by the empirical work points at knowledge being (at least) partly nonrivalrous (and thus giving rise to spillovers and possibly growth scale effects).

Gort (1962)

In his seminal work on diversification, Gort investigates, amongst others, the direction of diversification. He finds that firms are more prone to diversify into industries (four digit SIC-codes)[51] that are characterized by a high ratio of technical personnel where he uses the ratio of technical personnel to proxy for the ratio of R&D expenditures divided by sales on which concrete data is missing. Gort (1962) sees the idea that a high ratio of technical personnel signals good opportunities for future growth as a possible explanation for this finding. He also finds that firms that are characterized by a high ratio of technical personnel (relatively large R&D efforts) are more prone to diversify (see also Gort 1969). Finally, Gort finds that larger firms are more diversified as measured by four digit SIC-codes (see also footnote 49).

Scherer (1965)

Scherer examines the relations between R&D and a whole range of variables including diversification and firm size. He finds that the level of diversification (within a certain period, e.g. a year) is significant in explaining observed variance both in total patent output (within a certain period, e.g. a year) divided by total sales (within a certain period, e.g. a year) and in R&D employment (within a certain period, e.g. a year) divided by total sales (within a certain period, e.g. a year). The last finding of course denotes a positive relation between diversification and R&D intensity. What is more, Scherer also finds a strong (linear) relation between patent output and the number of R&D employees which means that the finding of a

[51] SIC is an abbreviation of Standard Industrial Classification.

positive relation between diversification and patent output divided by total sales also points to a positive relation between the level of diversification and R&D intensity.[52]

Grabowski (1968)

Grabowski investigates the relation between R&D intensity (R&D expenditures divided by total sales) and diversification (as measured by five digit SIC codes) for the chemical, drugs and petroleum industries. He finds a highly statistically significant positive relation between R&D intensity and a firm's level of diversification (the extent to which a firm is diversified) for the chemical and drugs industries but not for the petroleum industry, which is also the least R&D intensive industry of the three. This relation holds even though a measure of the available financial resources, which could proxy for size, has also been concluded as an explanatory variable in his estimating equation. Because diversification and the availability of financial resources are expected to be strongly related to firm size, Grabowski also estimates the relation between R&D intensity and size for the chemical and drugs industries. He finds a positive relation between size and R&D intensity for the chemical industry and a hump-shaped relation between these variables for the drugs industry.

Teece (1980)

In the appendix to his paper on economies of scope and the scope of the enterprise, Teece investigates the extent to which economies of scope played a role in the diversification of the petroleum industry into alternate [sic] fuels. He finds that a firm's level of diversification is positively related to its knowledge stock measured by its cumulative R&D expenditures. This relation holds even though cash flow, which might serve as a proxy for size, is included as an explanatory variable as well.

Link and Long (1981)

Link and Long investigate the relation between a firm's basic R&D intensity (with basic R&D defined as "original investigation for the advancement of scientific knowledge that does not have a specific commercial objective" (p. 105)) and its level of diversification. They find that more diversified firms, are more prone to conduct basic R&D (the more diversified a firm is,

[52] Scherer (1965) does not examine the relation between R&D employment and patent output to study economies of scale in the R&D process, but rather data limitations force him to make use of both R&D employment and patent output as measures indicating R&D efforts. Note however, that in principal the existence of (dis-)economies of scale in the R&D process makes patent output, or similarly, the level of productivity and productivity growth, unsuitable measures of R&D efforts (cf. Nelson 1990).

the higher also its basic R&D intensity is), even when other important determinants like type of industry (low-tech vs. high-tech) and firm size are controlled for.

Lemelin (1982)

Lemelin studies patterns of diversification by looking at pairs of origin-target industries. He uses a regression equation in which the probability of diversification is modelled as a function of various firm/industry characteristics and finds that more diversification takes place between industries that are what he calls 'science-based industries', i.e. industries that try to exploit science-based research.

Scherer (1983)

Scherer investigates variables that influence the patent output of firms. As would be expected, he finds a strong relation between the number of patents and R&D expenditures (see also Scherer 1965). The best fit was provided when besides a linear term also a quadratic term ($R\&D^2$) was incorporated as explanatory variable. The sign of the quadratic coefficient varied over different industries but in general there were no significant departures from constant returns to scale (74 out of 124 cases), and the cases that did depart generally had diminishing returns (31 out of 124). As additional variables, Scherer includes, amongst others, the level of diversification (as measured by a Herfindahl index). He finds that diversification stimulates patenting. That is, when R&D expenditures are accounted for, more diversified firms have a higher propensity to patent.

MacDonald (1985)

Patterns of diversification and the directions of diversification are examined by MacDonald. He finds that industries that are relatively R&D intensive both attract relatively much inbound diversification and generate relatively much outbound diversification. In addition, firms active in R&D intensive industries are more likely to diversify. When these firms diversify, they are most likely to enter industries similar to their origin industries. That is, they are more likely to enter R&D intensive industries.

Montgomery and Hariharan (1991)

Montgomery and Hariharan investigate the probability that a firm will diversify into a certain industry as a function of, amongst others, industry variables and firm-specific variables. For industries, they find that high R&D intensity industries attract inbound diversification.

Similarly, firms with large (past) R&D expenditures are more likely to diversify, as are firms with large (past) advertising expenditures[53]. Finally, firms are more prone to diversify into industries that are similar to their origin industries, e.g. industries that have the same R&D intensity.

3.3.2 R&D and firm size

Several studies have investigated the relations between R&D and diversification, e.g. Scherer (1965), Grabowski (1968), Soete (1979), Bound, Cummins, Griliches, Hall and Jaffe (1984), Cohen, Levin and Mowery (1987), Cohen and Klepper (1996a), and Cohen and Klepper (1996b). The results of these studies regarding the relations between R&D and firm size and other possibly interesting findings are summarized below except for the study by Grabowski (1968), which has already been discussed in the previous paragraph.
As in the previous paragraph, before starting the actual discussion of the available evidence, the expected relation between R&D and firm size is discussed. The concrete hypothesis regarding this relation follows from the idea formulated above, that size can be used to exploit the economies of scale and scope that knowledge's nonrivalrous nature gives rise to; R&D provides firms with the incentive to increase their size (in an attempt to make the widest use of its knowledge stock and thus exploit the economies of scale and scope associated with knowledge's nonrivalrousness).[54]
Again, the empirical work is used to establish whether this hypothesis is correct or incorrect which in turn will provide information on one of the basic questions, namely how knowledge, R&D and spillovers, i.e. knowledge's nonrivalrousness, affects the size of firms.

Scherer (1965)

As stated, Scherer (1965) examines the relations between R&D and, amongst others, firm size. He finds that smaller firms are responsible for a larger share of R&D efforts than could be expected from their share in total output. Scherer also finds that total patent output

[53] Again here advertising and R&D have to some extent similar effects; see also Megna and Mueller and footnote 44.
[54] One is inclined to associate diversification with economies of scope and firm size with economies of scale. Often, however, the difference between the two types of economies is small and artificial. In fact, when a firm can use the unique knowledge it has about producing TV's to build computer monitors as well there are economies of scope but at a more aggregated level, i.e. the level of producing consumer electronics, there are economies of scale. Because, in light of the subject of this thesis, not the economies itself are of interest here, but rather their common source, namely knowledge's nonrivalrousness, in this thesis diversification and size are linked to both type of economies.

increases more than proportional with firm size for smaller firms but less than proportional for larger firms, i.e. there is a hump-shaped relation. This finding in itself of course does not shed any light on how size influences the role knowledge, R&D and spillovers, play at the firm level, but, as mentioned above, Scherer also finds a strong (linear) relation between patent output and the number of R&D employees. This finding can thus be seen as conveying that R&D effort increases more than proportional with firm size for smaller firms, but less than proportional for larger firms.

Soete (1979)

Soete studies the relation between R&D expenditures by firms and their size. In contrast with Scherer (1965), he finds that in his sample as a whole R&D expenditure increases more than proportional with firm size. At the industry level, however, the results are less clear cut. Adding a quadratic term to the regression equations shows that in some industries R&D expenditures increase more than proportional with firm size, while in others R&D expenditures increase less than proportional. In the remaining industries, the quadratic term was not significant but the linear term was, indicating that in these industries R&D expenditures increase proportional to R&D.

Link (1980)

Link examines the relation between firm size and the rate of return to R&D expenditures. He finds that in his sample of firms from the chemicals and allied-products industry (SIC 28) the rate of return to R&D expenditures increases with firm size. The estimated rate of return is approximately 30% for relatively small firms, and 78% for relatively large firms. However, beyond a certain firm size there are no further benefits to be reaped from increasing the size of the firm.[55]

Bound, Cummins, Griliches, Hall and Jaffe (1984)

Bound et al. investigate, amongst others, the relation between firm size, as measured by total sales, and R&D intensity. They find an elasticity of R&D with respect to sales of close to unity. They have also found a U-shaped relation between firm size and R&D intensity,

[55] Similar, to Scherer's (1965) finding regarding firm size and patent output, the relation between the rate of return and firm size does not shed much light on how size influences the role knowledge, R&D and spillovers, play at the firm level. It seems logical, however, that a firm would increase its size until it reaches its profit maximizing size, which when based solely on returns to R&D investment is large.

however. That is, both small and very large firms are relatively R&D intensive and this is in contrast with the hump-shaped relation found by Scherer (1965) (see also Soete 1979).

Cohen, Levin and Mowery (1987)

Cohen, Levin and Mowery present several interesting findings regarding the relation between size and R&D intensity. They find that larger firms are more likely to conduct R&D (or at least report the fact that they do). Using dummy variables to control for industry effects, the authors find no relationship between R&D intensity and size at the business-unit level, however. At the firm level, in contrast, size does have a statistically significant though almost negligible effect on R&D intensity and this effect vanishes when some outliers are removed from the sample. Rather than size, Cohen, Levin and Mowery find that, fixed industry effects can explain a large portion, some 50%, of the variance in R&D intensity.

Cohen and Klepper (1996a and 1996b)

The studies of Cohen and Klepper seem out of place in this discussion of empirical work that has investigated the relation between R&D and firm size. However, these studies have investigated some interesting relations closely related to the hypothesis that R&D provides firm with an incentive to increase their size.

In their first study, Cohen and Klepper (1996a), investigate how firm size influences the allocation of R&D expenditures over process and product innovation respectively. They hypothesize that (i) for larger firms process innovation is more profitable than it is for smaller firms, because larger firms can apply the fruits of this type of R&D to greater output than smaller firms can, and that (ii) for product innovation size matters less because the fruits of this type of innovation can be sold in the market more easily, i.e. they are more saleable. In a regression estimate of the shares of process and product innovation as a function of firm size (and other moderating variables), Cohen and Klepper (1996a) subsequently find empirical support for their hypotheses. Overall, they find their results to be supportive of the idea that large firms can spread the costs of R&D over more products (see also Cohen and Klepper 1996b).

In their second study (Cohen and Klepper 1996b), they construct a model of the relation between R&D and firm size that uses the idea that firms can spread the costs of R&D over a large amount of products and find that the model and idea is supported by empirical evidence. The most elegant and convincing supporting finding is that the positive (proportional) relation between R&D and firm size is weaker for firms active in industries in which the fruits of

R&D are more saleable. That is, when there is less need for the firm to exploit the new knowledge itself (cf. Teece 1980 and 1982) firm size is less important for the profitable conduct of R&D (see also Cohen and Klepper 1996a).

3.3.3 Concluding remarks

The discussion of papers that have investigated the relation between R&D and diversification and R&D and firm size in the previous two paragraphs has presented several interesting findings.

To investigate how these findings relate to the influence knowledge, R&D and spillovers, have on diversification by firms and firm size, use is made of the two hypotheses, formulated at the beginning of each paragraph:

1. R&D leads to diversification (because diversification is a way to exploit to economies of scale and scope associated with knowledge's nonrivalrousness),
2. R&D leads to increases in firm size (because in this way a firm tries to make the widest use of its knowledge stock and thus exploit the economies of scale and scope associated with knowledge's nonrivalrousness).

The first hypothesis is confirmed in the empirical work on the relation between R&D and diversification. Generally, firms with relatively much firm-specific knowledge as evidenced by (past) R&D efforts, diversify more than other firms, i.e. firms with relatively little firm-specific knowledge, do. Furthermore, diversification generally takes place into similar type of industries, e.g. industries with high R&D intensity, which appears to indicate that there is little (technological) distance between the origin and the target industries. This means that the diversifying firm's knowledge stock is relatively useful in the new industry, which is to be expected when the goal of diversification is to exploit economies of scope, which due to knowledge's nonrivalrousness are associated with the firm's knowledge stock.
The second hypothesis, on the other hand, cannot be confirmed. Despite the finding of the empirical work on the relation between R&D and firm size that R&D effort increases roughly proportional with firm size, these studies provide no evidence that indicates that R&D provides firms with the incentive to increase their size (in an attempt to make wider use of their knowledge stock). In addition, often, the finding of a proportional relation between R&D and firm size is taken to indicate the existence of a scale effect, in turn related to the nonrivalrous nature of knowledge. This, however, is not the case, because there is a strong

causality problem in the relation between R&D and size: does larger firm size lead to more R&D, or does R&D lead to a larger firm size (cf. Nickell 1996).[56] All this is not say that the findings of the empirical work on the R&D-size relation provide support for the idea that knowledge is not nonrivalrous and that consequently R&D provides no incentive for firms to increase their size either however. Rather the remarks regarding this relation simply reflect the fact that these studies have not investigated the proper relation needed to draw conclusions regarding the hypothesis that R&D leads to increases in firm size.[57]

Nonetheless, some other conclusions may be drawn for the empirical work on the relation between R&D and diversification and R&D and firm size has come up with more interesting findings besides the ones that can be related to the two hypotheses.

These findings are that:

1. more diversified firms have higher R&D intensities,
2. as mentioned, R&D efforts increase approximately proportional with firm size,
3. the probability of firms conducting R&D rises with their size (larger firms are more likely to conduct R&D than are smaller firms),
4. R&D is more profitable in larger firms,
5. large firms spread the costs of R&D over more output, i.e. cost spreading of R&D takes place.

The first of these findings is consistent with Nelson's diversification hypothesis (cf. Link and Long) which says that diversified firms have an advantage (relative to less diversified firms) in conducting (basic) R&D. After all, if some type of firm has an advantage in a certain activity, e.g. conducting R&D, this type of firm is logically expected to devote more effort to this activity than other firms. Further support for Nelson's (1959) diversification hypothesis can be found in Scherer's (1983) finding that diversification stimulates patenting. This finding can be seen to signal that diversified firms have more opportunities to exploit the fruits of R&D profitably and are consequently more willing to incur the costs associated with a patent application (cf. Scherer 1983).

---

[56] The causality problem in the R&D firm size relation is a thorny issue that surfaces again below when other findings of the studies discussed in the previous two paragraphs are discussed, and there also this issue is elaborated on.

[57] Two studies that may be supportive of the hypothesis that R&D leads to size increases are those of Mansfield (1962) and Hall (1987) discussed in the first section of this chapter and Appendix B. Mansfield finds that successful innovators grow, on average, twice as fast as other firms ceteris paribus. Hall finds that an increase in R&D expenditures of 2 million dollars in 1976 is associated with a 1 percent increase in the annual growth rate for the period 1976-1979 and that the same effect is achieved with a 4 million dollar increase in investment. Whether the resulting increase in firm size is the result of deliberate action of the firm or just the result of increased demand for their products is subject to debate however.

The other four findings are consistent (or at least not inconsistent) with the so-called 'Schumpeter hypothesis' or 'Schumpeterian hypothesis' (e.g. Link 1980 and Nelson 1990). This hypothesis originates from Schumpeter's (1942 | 1944) remark that "… the large-scale establishment or unit of control must be accepted as […] the most powerful engine of that progress and in particular of the long-run expansion of total output ..." (p. 106). Other researchers subsequently have refined Schumpeter's remark to refer to the idea that larger firms are better suited to perform R&D than smaller firms because they are able to, amongst others, (i) apply the fruits of R&D to relatively more output and/or spread the costs of R&D over a larger amount of products, and (ii) initiate and maintain more R&D projects and subsequently minimize the risk that is associated with a stochastic process like R&D (cf. Arrow 1962, Nelson, Peck and Kalachek 1967, and Nordhaus 1969).[58] Hence, just like Nelson's diversification hypothesis, above the Schumpeter hypothesis is confirmed in the empirical work discussed above.

However, what is needed to wash away the disappointing results regarding the hypothesis that R&D stimulates size increases, is not conformation of either Nelson's diversification hypothesis or the Schumpeter hypothesis. Rather, the question that is of interest is whether these five findings reveal something (anything) about knowledge's nonrivalrousness. Unfortunately this is only unambiguously the case for the fifth finding, i.e. the finding that cost spreading of R&D takes place, for this can only occur if knowledge, in turn created through R&D effort, is at least partly nonrivalrous. Drawing conclusions regarding knowledge and its nonrivalrousness from the other four findings in contrast is not possible because, factors like advertising expenditure, size in actual product markets (as opposed to firm size), R&D effort etc., are mutually related: for example R&D increases demand for a firm's products and hence its size which through cost-spreading stimulates R&D. Hence, in much the same way as with the relation between the proportional relation between firm size and R&D discussed above there is a causality problem limiting the possibility of drawing conclusions regarding the nonrivalrous nature of knowledge and its nonrivalrousness from these (four other) findings. Without additional information on what can best be broadly

[58] There is some controversy on the correct interpretation of Schumpeter's remarks on large firms and innovation and consequently on the nature of the 'true' Schumpeterian hypothesis however (cf. Fisher and Temin 1973, McNulty 1974, and Kamien and Schwartz 1982). Rather than number of employees or total sales, size in the Schumpeterian interpretation is said to convey some sort of market (monopoly) power and a test of the Schumpeterian hypothesis would thus be a test of the incentive effect a measure of market power has on R&D effort. Oft cited in this respect is Schumpeter's remark that "… size is neither necessary nor sufficient for it." (p. 101). Nevertheless, the two arguments in support of the idea that larger firms are better suited to perform R&D than smaller firms have some strength in their own right regardless of whether they convey the intent of Schumpeter or not and the idea will continue to be referred to as the Schumpeter hypothesis.

referred to as the market situation (e.g. market structure) the above-presented findings (including the fact that R&D efforts increase approximately proportional with firm size) can be judged as being consistent with either of two extreme cases: knowledge being completely nonrivalrous or completely rivalrous.[59]

The preliminary conclusion following the evidence of R&D providing and incentive for firms to diversify that knowledge is at least partly nonrivalrous holds its grounds. The real question consequently remains; why is there no scale effect in growth even though knowledge is at least partly nonrivalrous?
The next section tries to answer this question by investigating whether knowledge and the exploitation of the economies of scale and scope it gives rise to have any real effects on performance by firms, e.g. on profitability.

3.4 Diversification and firm performance

The previous section began by arguing that knowledge's nonrivalrous nature gives rise to economies of scale and/or scope: a firm is able to achieve lower average and marginal costs by applying its resources, e.g. knowledge, to a larger number of products (economies of scale) and/or to a wider range of products (economies of scope). The idea that firms can use diversification to exploit these economies of scale and (especially) scope (cf. Penrose and Teece) was supported by empirical evidence. The studies discussed in the previous section have investigated the relations between R&D and diversification and have shown that more (past) R&D efforts increase diversification by firms and that more diversified firms conduct more R&D. The force underlying these relations is of course firms striving for profits and consequently exploiting economies of scale and scope resulting from knowledge's nonrivalrousness through diversification. The implication of this simple idea is that diversification, and ultimately knowledge and the extent to which it is nonrivalrous, may determine a firm's performance, e.g. its profitability. Stated differently, diversified firms are expected to perform better than their (relatively) undiversified peers do, because they are able to exploit economies of scope and scale resulting from, amongst others, the nonrivalrous nature of knowledge. Studying the effect diversification has on firm performance (e.g.

[59] In the empirical work on the relation between R&D and diversification used to test the first hypothesis that R&D leads to diversification this problem of an incompletely specified model is of course also present, but severely limited because the relation that is apparent in this work is causal, running from R&D to diversification.

profitability) subsequently seems a promising area of research for assessing the role knowledge, R&D and spillovers, play at the firm level. The relation between diversification and performance is widely studied and this section will discuss some of the work done on this so-called diversification-performance nexus.

The section starts with a discussion of some of the difficulties that arise when using the work done on the diversification-performance nexus to analyze the issue of knowledge, R&D and spillovers, at the firm level (Paragraph 3.4.1). Paragraph 3.4.2 subsequently discusses some of the empirical work done on the diversification-performance nexus. Some proposed theoretical models consistent with the empirical evidence on the relation between diversification and performance (e.g. profitability) are discussed in Paragraph 3.4.3. The idea is of course that this theoretical work may have treated knowledge, R&D and spillovers, differently than the literature on models of growth without scale effect has, and consequently provide new ways of credibly modelling growth without scale effects. The section ends with some concluding remarks, which are presented in Paragraph 3.4.4.

3.4.1 Motives for diversification

The idea behind this paragraph is that knowing the magnitude and nature of the performance consequences of diversification are not open for straightforward interpretation regarding the role knowledge, R&D and spillovers, play in firms. That is, knowing that diversification level does have a positive (or negative) effect of a certain magnitude on firm performance is not enough to draw robust conclusions on the extent of the economies of scale and scope associated with knowledge's nonrivalrous nature. The diversification-performance nexus is simply too complex to make such a bold statement on so (relatively) little information.

The two most prominent difficulties that arise when trying to use research on the relation between diversification and performance to draw conclusions on economies of scale and scope and consequently the role knowledge, R&D and spillovers, play in firms are that:

1. there are more reasons for firms to diversify than the exploitation of economies of scale of scope,
2. it is hard to discern whether a firm tries to exploit economies of scale and scope that are related to capital and human capital or those that are related to the firm's knowledge stock and the nonrivalrousness of knowledge.

Ad 1

Regarding the first difficulty, the multitude of motives for diversification, two broad categories of reasons why firms diversify can be discerned.[60] The first category (1) contains the motives that are consistent with the standard assumption in economics that a firm's one and only goal is to maximize profits. In this view, the firm's management pursues a strategy of diversification to achieve its goal of profit maximization. The second category (2) acknowledges that in the determination of a firm's strategy, agency problems play a large role. Top-level executives may not act in the firm's best interest, maximize profits, but rather use the firm to further their own interests.

1. The motives for diversification that can be shared under the first category are (i) resource synergies, that is economies of scale and/or scope, (ii) market power, (iii) financial synergies, e.g. the benefit of an internal capital market, and (iv) entering new promising industries (greener pastures).
    - Of all the motives for diversification, resource synergies are of course of most interest in this thesis in general and in this section in particular. Resource synergies are closely related to economies of scale and/or scope which in turn are related to the increasing returns the nonrivalrous nature of knowledge gives rise to. Diversification may subsequently be used to exploit these economies (see also Section 3.3).
    - Market power can be defined as the ability of a firm to control the price, quantity and/or the characteristics, of the products the firm sells in a given market. A firm that has market power can use this power to increase its profitability relative to a firm that operates equally efficient (Montgomery 1994, p. 165). A (large) diversified firm can achieve market power and increase its profitability through cross-subsidization (predatory pricing), mutual forbearance and reciprocal buying (Montgomery, p. 165).
    - Unsurprisingly, financial synergies are very similar to resource synergies. However, where resource synergies are related to the physical operation of the different segments within a diversified firm and the cross-connections between them, financial synergies are related to intangible aspects of the firm. For diversified firms financial synergies come in several forms, e.g. the benefit of a (large) internal capital market, tax advantages, and co-insurance (cf. Lewellen 1971).
    - Through diversification, firms can enter promising new industries. A firm may be

[60] Montgomery (1994) provides and excellent discussion of motives for diversification, including most of the ones discussed here.

active in a declining industry with poor prospects and diversify towards greener pastures, i.e. diversify into a young industry with good prospects for future growth and profits.

2. The second broad type of reasons why firms diversify contains all the ways in which the firm's top-level executives can further their own interests (rather than the firm's) through diversification. Agency theory predicts that when management's stake in the firm is limited, the decisions taken by the management will not always be in the best interest of the firm but rather in the interest of management itself. Jensen (1986) argues that the agency problem is particularly severe when it regards the so-called free cash flow, the "[…] cash flow in excess of that required to fund all projects that have positive net present values when discounted at the relevant cost of capital." (p. 323). Shareholders would like to see the free cash flow being paid out to them in the form of dividend, while the firm's management can use it to further their own benefits, e.g. through firm growth. Mueller (1969) sees firm growth as one of the goals of managers because, amongst others, their salaries and bonuses are tied to the size of the firm, and because the power and prestige managers derive from their occupation is directly related to the size of the firm (p. 644). Similarly, Shleifer and Vishny (1989) describe how managers can use investment projects, e.g. diversifying into a new line of business financed from the free cash flow, to increase demand for their particular skills (managerial entrenchment).

Thus, diversification may contribute to a firm's performance for several reasons, including the exploitation of economies of scale and/or scope resulting from knowledge's nonrivalrousness. However, its effect on performance may also be negative when the firm's management uses it to increase their own well-being and do not use diversification's potential for performance improvement. As stated above, separating the effect of knowledge and the concomitant economies of scale and/or scope from the effects of other factors is subsequently difficult.

Ad 2

The second difficulty, determining whether diversification is aimed at the exploitation of economies of scale and scope that are related to capital and human capital or those that are related to the firm's knowledge stock and the nonrivalrous nature of knowledge, is also hard to solve. A possible solution would be to conduct a broad comparison of the characteristics of the origin industry/firm and the target industry/firm. Much in the same way as studies, like those by Lemelin, MacDonald and Montgomery and Hariharan, discussed above have described patterns of diversification or as Jaffe (1986) has defined a firm's technological

position, firms/industries can be viewed as a set of combined characteristics. Similarity in the characteristics of the origin and target industries/firms are expected to signal that a firm's current resource stock might be useful in the new industry (e.g. technological proximity; cf. Jaffe 1986) (see also Lemelin 1982, MacDonald 1985, and Montgomery and Hariharan 1991).[61] That is, similarity between origin and target industries/firms is a logical representation of the possibility of economies of scale and/or scope when both the origin and target type of activity are performed within a single firm. The extent to which knowledge, capital, and/or human capital, is/are the resource underlying the economies of scale and/or scope resulting form the diversification, should subsequently be assessed by analyzing the role each of these resources play in the similarities. This can be clarified by two simple examples. Imagine two countrywide operating retail chains, one selling books, and one selling shoes. When each of these two chains operates its own distribution network complete with warehouses and a fleet of trucks, it seems that these activities can be combined and capital can be saved (a scale economy related to capital). Similarly, imagine a firm that has created a revolutionary new way of welding steel plates together and uses this technique to produce cars. Since in a shipyard a lot of welding takes place as well, it seems logical to apply the technique in that industry as well (a scope economy related to knowledge). Analyzing each diversifying activity a firm undertakes in this way seems to be the most logical - and maybe even the only - way to distinguish between the several sources leading to economies of scale and/or scope. In addition, it may provide useful data on certain regularities underlying diversification (again see Lemelin, MacDonald, and Montgomery and Hariharan).

Having discussed the two most prominent difficulties, it is obvious that one must be very cautious when using research on the diversification-performance nexus to draw conclusions regarding the role knowledge, R&D and spillovers, play in firms. There are simply many factors that cloud the real contribution of economies of scale and scope associated with the nonrivalrousness of knowledge to firm performance.

[61] Note that it is implicitly assumed here that firms diversify into industries to exploit economies of scale and/or scope. However, when diversification is motivated by the will to find greener pastures or management trying to further their own interests, the extent of such economies may very well be limited. On the other hand, two industries/firms will always have some similarities, which in turn could signal possible usefulness of the firm's resource stock, i.e. possible economies of scale and/or scope.

3.4.2 Empirical work

Following the seminal contributions of Gort (1962) and Rumelt (1974), there are now literally hundreds of studies that have investigated the diversification-performance nexus and it is one of the most widely studied subjects in strategic management. Providing a complete overview of the findings of the literature is thus an impossible task.
Therefore, the first part (i) of this discussion of the empirical work on the diversification-performance nexus will be based on some survey articles which have summarized the first 20 to 30 years of the literature, namely Grant, Jammine and Thomas (1988), Ramanujam and Varadarajan (1989), Datta, Rajagopalan and Rasheed (1991), and Dess, Gupta, Hennart and Hill (1995). Examination of these overview articles and their most important results should provide a broad insight into the relation between diversification and performance.
In the second part (ii) subsequently recent work on the diversification performance nexus will be discussed, namely Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), Lins and Servaes (1999), Rajan, Servaes and Zingales (2000), Campa and Kedia (2002), and Graham, Lemmon and Wolf (2002). Finally (iii), at the end of this paragraph, two studies, Lichtenberg (1992) and Schoar (2002), will be discussed that have used productivity as their measure of performance. Using productivity as performance measure places these studies in a context that is closer to the study of economic growth, than are studies that have used performance measures like profitability, returns on equity, etc.

Early work

As stated, Grant, Jammine and Thomas, Ramanujam and Varadarajan, Datta, Rajagopalan and Rasheed and Dess et al., discuss the results of the first 20 to 30 years of the empirical work on the relation between diversification and performance. Partly because the articles examine roughly the same studies of the diversification-performance nexus, they tend to agree with each other and reach the same conclusions regarding the results of this 'early work'. Hence, here only the most important combined findings of these overview articles are discussed. A more detailed discussion of the individual studies is presented in Appendix D.1.
The most important finding of the studies by Grant, Jammine and Thomas, Ramanujam and Varadarajan, Datta, Rajagopalan and Rasheed and Dess et al., is that the early work investigating the diversification-performance nexus has provided no solid results, i.e. after 30 years of research still very little is known with certainty about the relation between diversification and performance. According to the authors of the surveys, this is largely due to

weaknesses in the research methods used in most of the work. They find that the most frequent and persistent weaknesses of the early work on the relation between diversification and performance are related to (i) the measures of diversification used, (ii) the measures of performance used, (iii) the lack of incorporation of possible moderating - and additional explanatory - variables, (iv) the cross-sectional nature of most research, (v) and the focus on variables at the corporate rather than at the business-unit level.

- Diversification has been measured using a broad range of variables. Widely used measures of diversification are the number of different two/three/four-digit SIC codes a firm is active in (a continuous measure), and Rumelt's scheme in which firms are categorized according to the nature of their diversification strategy (a categorical measure). Both these (type of) measures have the drawback of hiding possibly important information.
- Similar, to diversification, performance has been measured using various variables including accounting based measures like return on equity and market-based measures like the risk premium ratio. However, any measure of performance at a disaggregated level should include both risk and return in its operationalization. Not taking into account either of these leads to incorrect results regarding the performance consequences of diversification.
- Diversification is only one of many variables that are expected to have an effect on a firm's performance. Market power, R&D investment, industry structure etc., are all related to a firm's performance and, quite possibly, to the strategic decisions it makes.
- Most early work on the relation between diversification and performance has examined large samples using cross-sectional regression analyses Besides the fact that in these analyses important explanatory variables like industry effects are often missing, the results found are only representative for a single point in time. Diversification on the other hand, may very well have effects that continue for several periods which means that the results cannot be generalized over time.
- Finally, the studies in the first 30 years of research on the diversification-performance nexus have investigated the effects of diversification at the corporate level rather than at the business-unit level where diversification actually takes place. The effect of diversification on performance, positive or negative, is subsequently moderated by the performance of other business units in the firm.

Regarding future work Grant, Jammine and Thomas, Ramanujam and Varadarajan, Datta, Rajagopalan and Rasheed, and Dess et al. state that it should take a different approach than hitherto and address the theoretical and methodological shortcomings of the early work.

Recent work

As stated, Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), Lins and Servaes (1999), Rajan, Servaes and Zingales (2000), Campa and Kedia (2002) and Graham, Lemmon and Wolf (2002), are some recent contributions to the literature on the diversification-performance nexus. They generally have acted upon the devastating critique of Grant, Jammine and Thomas, Ramanujam and Varadarajan, Datta, Rajagopalan and Rasheed and Dess et al., on earlier empirical work. These papers emanating mainly from the finance literature have addressed some of the weaknesses of earlier work on the relation between diversification and performance emanating mainly from the strategic management literature. The findings of the studies by Lang and Stulz (LS94), Berger and Ofek (BO), Servaes (Servaes), Lins and Servaes (LS99), Rajan, Servaes and Zingales (RSZ), Campa and Kedia (CK) and Graham, Lemmon and Wolf (GLW), with respect to the performance effect of diversification (Diversification discount: yes or no) are summarized in Table 3. In addition, Table 3 provides information on the measure of performance used (Performance measure), the measure of diversification used (Diversification measure), whether the direction of causality, from diversification to performance or from performance to diversification, has been investigated (Causality), the factors, variables etc., used in case the results are tested for robustness (Robustness), and other interesting findings (besides the performance effect of diversification) (Other). A more detailed discussion of the various studies can be found in Appendix D.2.

The papers summarized in Table 3 of course present only a cross-section of the work done on the diversification-performance nexus. At first sight however, their evidence on the performance effects of diversification is clear enough to allow drawing the conclusion that diversification has a negative effect on firm performance, measured by excess value (constructed from the ratio of Capital (to Assets, Sales and/or Earnings) or Tobin's q (cf. Tobin 1969). That is, diversified firms sell in the market at a discount relative to their less diversified peers, and their thus is a diversification discount. However, in this conclusion some nuances must be made when the other findings of the studies by Lang and Stulz, Berger and Ofek, Servaes, Lins and Servaes, Rajan, Servaes and Zingales, Campa and Kedia and Graham, Lemmon and Wolf, are taken into consideration. In line with the recommendations

Table 3: Main features of some recent work on the diversification-performance nexus.

Study	Diversification discount	Performance measure	Diversification measure	Causality	Robustness	Other
LS94	Yes, for several measures of diversification	Tobin's q	Two Herfindahl-type indices and the number of individual segments (four-digit SIC-codes)	Diversifying firms have lower q, but their q is equal to those of (single-unit) firms in the same industry	Industry effects and size, capital market access and R&D intensity as explanatory variables	-
BO	Yes, for several differently constructed performance measures, and averaging 13-15%	Excess value constructed from the ratio of Capital (the sum of book value of debt and market value of equity) to Assets, Sales and/or Earnings	The number of individual segments (four-digit SIC-codes) (1 = undiversified; > 1 = diversified)	-	-	Overinvestment and cross-subsidization cannot explain the discount
Servaes	A positive discount for the period 1961-1970 but a negative discount for the period 1970-1973	Tobin's q	A dummy variable which takes the value 1 when the firm is active in more than one industry (two-digit SIC-codes)	Firms have lower q because they are diversified, not because firms with lower q diversify	Industry effects and size, capital market access and R&D intensity as explanatory variables	Firms, especially those with high insider ownership increase their diversity in response to lower costs (of diversification)
LS99	A positive discount for the U.K. (15%) and Japan (10%), but a zero discount for Germany	Excess value (cf. Berger and Ofek)	A dummy variable (cf. Servaes 1996 (Servaes))	-	Industry effects	Insider ownership can only partially explain the observed cross-country differences
RSZ	Yes, and the discount increases with increasing diversity	Tobin's q	Diversity measured as dispersion in weighted opportunities which are measured by Tobin's q			Higher diversity leads to more inefficient investment
CK	Yes, but the discount decreases (to a premium) when the endogeneity of diversification is controlled for	Excess value (cf. Berger and Ofek)	A dummy variable (cf. Servaes 1996 (Servaes))	See robustness	Fixed-firm effects, endogenizing the decision to diversify, and self-selection	-
GLW	Yes, but about one half of the discount is due to the acquisition of already discounted firms	Excess value (cf. Berger and Ofek)	The number of individual segments (four-digit SIC-codes) (1 = undiversified; > 1 = diversified)		Pre-acquisition performance of target firms	Performance differences between related and unrelated acquisitions

Source: Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), Lins and Servaes (1999), Rajan, Servaes and Zingales (2000), Campa and Kedia (2002), and Graham, Lemmon and Wolf (2002).

of Grant, Jammine and Thomas, Ramanujam and Varadarajan, Datta, Rajagopalan and Rasheed and Dess et al., the recent work on the diversification-performance nexus addresses some of the shortcomings of the earlier papers. Besides the use of excess value and/or Tobin's q as measures of firm performance, in which both risk and return are represented, they also include additional explanatory variables for instance. It is when these additional variables are included, that the performance effect of diversification, i.e. is there a diversification discount or even a diversification premium, becomes less clear. Most studies find that other variables besides diversification can at least partly explain the negative performance effect previously attributed to diversification. Lins and Servaes find that the performance consequences of diversification differ between countries and the study by Campa and Kedia even find that poor circumstances (e.g. declining market) lead firms to diversify after which they actually perform better, a negative diversification discount or a diversification premium. Hence, the results on the relation between diversification and performance are still mixed, though less mixed than the findings of the early work but a safe conclusion would be that diversification has no - or a slightly negative - effect on performance. What is more, the recent work on the diversification-performance nexus discussed here has not addressed all shortcomings of the early work. First of all, (i) the studies still lack a time component, i.e. they are still cross-sectional in nature, although the use of partly market-based measures of performance like excess value and Tobin's q, does bring a bit of time dimension in the performance effects of diversification. Second, (ii) they still use inadequate measures of diversification (number of two/four-digit SIC codes).

The most logical conclusion from the empirical work on the diversification-performance nexus discussed above is that any benefits of economies of scale and scope resulting from knowledge's nonrivalrousness, are (more than) offset by the costs associated with being active in more than one line of business (diseconomies of scale and scope), e.g. dilution of knowledge (R&D) or loss of control (cf. Adams and Jaffe 1996 and Williamson 1967). Of course, economies of scale and/or scope can also be absent all together, leaving diseconomies of scale and scope as the only performance effect of diversification.

To examine the issue further, here, at the end of this paragraph, the performance consequences will be discussed outside a financial setting. Rather than excess value or Tobin's q, Lichtenberg and Schoar have used productivity as their measure of performance. Productivity in turn has a much more logical relation to the pervading issue of economic growth.

Lichtenberg (1992)

Lichtenberg views TFP as a purer measure of technical efficiency than for example Tobin's q. He subsequently investigates TFP at the plant level as a function of (i) the total number of four-digit SIC manufacturing industries a firm is active in (NINDS), (ii) the fraction of the firm's plants that operate in the same industry as this plant (SAMEIND), (iii) the total number of manufacturing plants the firm comprises (NPLANTS), (iv) a dummy variable equal to one if the firm operates only one plant (SINGLE), (v) the fraction of the firm's employees working in auxiliary (central) establishments (e.g. in a head-office or in a R&D laboratory) (AUXSHARE), and (vi) the size of the plant as measured by its total number of employees (PLANTEMP). Lichtenberg begins with determining the correlation between his explanatory variables and finds, unsurprisingly, that (1) NINDS and NPLANTS are strongly positively related[62], (2) SAMEIND and NPLANTS are strongly negatively related, and (3) SAMEIND and NINDS are strongly negatively related. He continues with estimating the regression equation for TFP and finds some remarkable results. The natural logarithm of NINDS (diversification) has a statistically highly significant negative impact on plant-level TFP as does the dummy that indicates that a firm consists of only one plant. SAMEIND contributes only statistically significant to TFP when NINDS is not included in the regression equation: NINDS dominates SAMEIND with which it is has a strong negative correlation. The natural logarithm of NPLANTS, AUXSHARE and the natural logarithm of PLANTEMP, all have a highly significant positive impact on the TFP level of the plant.[63] What is more, the coefficients for the natural logarithm of NINDS and the natural logarithm of NPLANTS are approximately equal in magnitude (though the opposite sign). It thus seems that TFP at the plant level does not depend on NINDS, SAMEIND, and NPLANTS separately, but rather on the number of industries divided by the total number of plants (NINDS/NPLANTS). Since SINGLE is found to have a statistically highly significant negative relation to productivity at the plant level, it would appear that plant-level productivity is negatively related to the number of industries per plant operated by the firms, and positively related to the number of plants per industry.

[62] This finding can be seen to indicate that more diversified firms are also larger, as measured by for instance total firm employment. After all, total employment is equal to the sum of all plant level employment and it seems unlikely that in diversified firms average plant employment is small enough to offset the effect of a larger number of plants.

[63] The finding that productivity increases with size (PLANTEMP) is consistent with the findings of Aw discussed in Section 3.1

Schoar (2002)

Schoar studies the effect diversification has on productivity from two different viewpoints, a static cross-sectional one and a dynamic longitudinal one. In the static analysis, she finds that plants in diversified firms are on average 7% more productive than similar plants belonging to undiversified firms active in the same industry.[64] Schoar further finds a life-cycle effect, older plants are less productive than younger plants, while size, measured both at the firm level and at the segment level, has a mixed effect on TFP. In the dynamic analysis, the picture of the effect of diversification on TFP changes. This type of analysis reveals that increases in diversification lead to decreases in firms' overall levels of productivity. This is the result of a decline in productivity in firms' incumbent plants of about 2%, which more than offsets the 3% productivity increase in the newly acquired firms. Thus, the dynamic effect of diversification is negative, while in the cross-section diversified firms remain more productive than their undiversified peers are. Schoar finds that misguided attention of management for the newly required plant ('new toy') may be partly responsible for the observed differences in the productivity-consequences of diversification for incumbent and acquired plants respectively. She continues with performing some robustness checks of the findings and finds that neither centralization of relatively unproductive tasks like administration in central non-manufacturing establishments nor transfer pricing can explain the results. Finally, the results are related to more commonly used performance measures, e.g. excess value. Schoar finds that a firm's excess value and its level of TFP are closely related but that, although diversified firms are more productive, they still sell at a discount.

3.4.3 Theoretical work

Having discussed some empirical work on the diversification-performance nexus, now some theoretical work on the nexus will be discussed in the hope that it provides additional insight on the role knowledge, R&D and spillovers, play at the firm level.

The number of models of the performance consequences of diversification stands in sharp contrast with the amount of empirical work on this subject. Undoubtedly, this results from the fact that it is far from clear what actually would require explanation, a diversification discount, a diversification premium, costs associated with diversification, or lacking benefits of diversification. This discussion of a relatively modest literature focuses on two theoretical

[64] This finding is of course at odds with that of Lichtenberg who finds that the total number of four-digit SIC manufacturing industries a firm is active in has a negative effect on plant-level productivity.

explanations, the incomplete contract theory explanation offered by Rotemberg and Saloner (1994), and the search process explanation developed by Matsusaka (2001).

Rotemberg and Saloner (1994)

Rotemberg and Saloner (1994) is one of the first papers to develop a model able to explain how the benefits of diversification, e.g. exploitation of economies of scale and scope, in spite their logical existence can ultimately not be reaped. In their model being active in a single (narrow) line of business is ultimately most profitable even though it means forgoing apparently profitable opportunities. In the Rotemberg-Saloner model worker effort in the R&D process is unobservable so that firms, to overcome the incentive problem this creates and deter shirking, make up a contract with their R&D personnel that promises them a payment conditional upon success (incomplete contracts). R&D personnel only gets paid if their ideas (e.g. a new product or a process improvement) are actually implemented. This effectively turns around the incentive problem from the R&D employees to the firm, for now the firm will check whether the benefits of implementing an idea exceed the costs (the payment to the employee). If this is not the case, the firm will not implement the idea and will not pay the employee. With synergies, e.g. economies of scope, between the different lines of business (business units) of the firm, this problem of strategic behavior on the part of the firm worsens. In this case, R&D personnel active in segment i might not be paid at all for their ideas (and effort) because their ideas are not implemented because the firm finds it cheaper to implements ideas developed by R&D personnel in segment j (for segment j) (spillover), even though these ideas are probably not as productive in segment i as original segment i ideas. That is, the amount the firm saves on the cost of R&D conducted by employees in segment i, more than offsets the negative productivity differential associated with the implementation of less suitable ideas from segment j. However, the strategic behavior of the firm causes R&D personnel in segment i to stop searching for ideas all together ultimately causing a deterioration in the firm's profitability. Hence, in Rotemberg and Saloner's model diversification leads to lower profitability and the solution to the incentive problems associated with incomplete contracts is for a firm to focus on a single 'core activity'.
Some critical remarks regarding their model must be made, however, especially concerning the way in which they model spillovers. First of all, (1) Rotemberg and Saloner model spillovers asymmetrically, i.e. ideas emanating from segment j can also be used in segment i, while no spillover mechanism exists through which ideas emanating from segment i can also be used in segment j. Second, (2) R&D efforts in one segment do not influence the

productivity of R&D in other segments. Hence, in the Rotemberg-Saloner model spillovers seem to be limited in an artificial an unearned way that resembles the assumptions the models of growth without scale effects discussed in the previous chapter make regarding the extent of spillovers, and when the assumptions regarding spillovers are relaxed the outcome of their model may be much more ambiguous. That is, modelling spillovers in a more realistic way may cause the result of the Rotemberg-Saloner model that focussing rather than diversifying is most profitable to hold no longer. These critical remarks, notwithstanding, the Rotemberg-Saloner model is quite insightful: the idea that rewarding R&D effort becomes harder with more lines of business and the existence of synergies remains, and it can be seen as a fruitful approach to modelling diseconomies of scale (cf. Williamson) even so as to provide a theoretical foundation for the dilution of R&D found by Adams and Jaffe.

Matsusaka (2001)

In Matsusaka's model diversification is the value maximizing strategy although focussing is generally efficient. The central idea is that a firm has 'organizational capabilities' (firm-specific knowledge / organizational capital / unique core competencies / etc.) that can be used profitably in different lines of business (cf. Penrose and Teece; see also the previous section). Matsusaka subsequently develops a fundamentally dynamic model in which firms continuously enter and exit lines of business in search for the best, most profitable, match for their firm-specific capabilities. This search process is fraught with uncertainty and upon entering a new line of business does not know for certain whether this line of business is a good match for its unique capabilities. When it is not a good match, the activity will be divested and the search for a good match continues. When the new line of business is a good match, the firm further specializes in this new line of business. The process of finding profitable matches for a firm's organizational capabilities continues however. The historical experience of a lot of firm follows the pattern described here, and many have actually changed their primary industry over the years. What is more, Matsusaka's search theory is able to explain observed negative performance effects of diversification. The diversification discount represents the fact that a large fraction of all firms has not yet found a good match for its organizational competencies. That is, diversified firms can (temporarily) sell at a discount although diversification is ultimately a value maximizing strategy.

3.4.4 Concluding remarks

This section has discussed papers that have studied the relation between diversification and firm performance in the hope that they would reveal something about the role knowledge, R&D and spillovers, play at the firm level. Unfortunately, the performance consequences of diversification are unclear, although diversification does appear to be associated with a slight negative performance effect. Combining the disappointing findings of the literature on the diversification-performance effect with the difficulty of determining the exact influence of knowledge's nonrivalrousness in the performance effect of diversification means that drawing conclusions on the existence and extent of economies of scale and/or scope resulting from knowledge's nonrivalrousness and in this way on the role knowledge, R&D and spillovers, play in firms, is very difficult. Thus, a modest conclusion is in place. Such a suitable conclusion would be that just like there is no scale effect in growth even though one would logically expect there to be one due to the nonrivalrous nature of knowledge, there are no obvious positive performance consequences of diversification even though one would logically expect a positive performance effect, resulting from the exploitation of economies of scale and/or scope resulting from, amongst others, knowledge's nonrivalrousness, through diversification.

The most important, and far less modest conclusion, however, is that despite the fact that many weaknesses in the research on the diversification-performance nexus were already pointed out some 15 years ago (cf. Grant, Jammine and Thomas), recent work still suffers from several shortcomings and that this is probably the main reason why the empirical work on the diversification-performance nexus has rendered so little concrete results.

In the three remarks presented below, fruitful directions for future work are put forward, two of which follow directly from persistent weaknesses/shortcomings of the work done on the relation between diversification and performance.

1. The most important weakness is the continued use of cross-sectional regression analysis with little interest for the longitudinal effects of diversification. That dynamic effects can be important and require further research is nonetheless shown by the search process theory of Matsusaka. How the time dimension can further be important when assessing the performance consequences of diversification is shown by a simple example. Imagine two firms, i and j, which when operating independently have annual profits of 100 each so

that their combined independent profits equal 200. In a simple cross-sectional analysis and using profits as the measure of performance, combining (merging) these two firms in a single firm (firm ij) shows a diversification discount when the profits of this new firm are below 200. Suppose, however, that the profitability of firm i, firm j and firm ij differs over time in the way depicted in figure 4 below.[65]

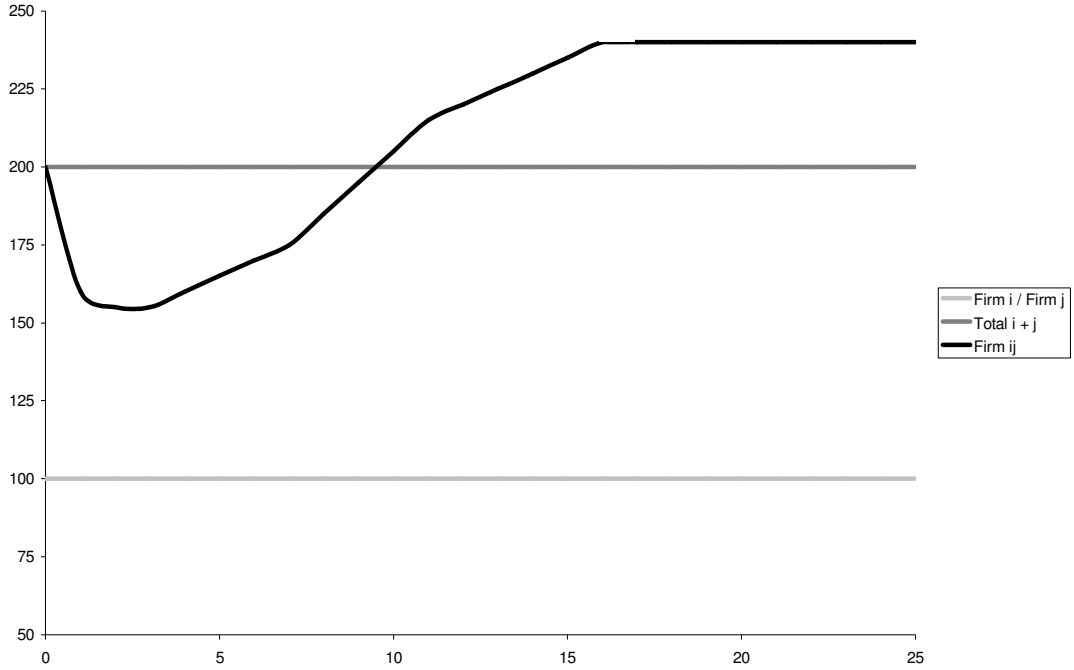


Figure 4, Profitability of independent firms i and j and merged firms i and j.

At time t = 0 the firms i and j are not yet combined into a single firm. At the time of the merger, t = 1, the profitability of the merged firms declines relative to the situation in which the firms would have remained independent. The idea is that merging two previously independent firms into one new firm is not straightforward and brings with it several costs. In fact, the profitability of the new merged firm does not reach the level of profitability of the combined independent firms until time t = 10. Hence, from time t = 1 until time t = 9, profitability as a measure of performance would indicate that diversification has negative performance consequences. However, when instead of

[65] The data belonging to Figure 4 are:

Time (t)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Firm i / Firm j	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Total i + j	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200
Firm ij	200	160	155	155	160	165	170	175	185	195	205	215	220	225	230	235	240	240	240	240	240	240	240	240	240	240

profitability a measure like Tobin's q, which incorporates future profitability (net present value), is used, the problem seems to be solved. Tobin's q is a market-based measure and hence in each period will the Net Present Value (NPV) of future profits will be reflected in Tobin's q, which consequently gives a much more accurate picture of the real performance effect of diversification than plain profitability does. However, only part of the problem of using cross-sectional analysis is mitigated by the use of Tobin's q or any measure incorporating NPV. The value of Tobin's q like profitability still depends on the time at which the analysis takes place. At time t = 1, the NPV of future profits differs significantly from the NPV of future profits at time t = 13. Hence, timing is still crucial. In addition, the NPV of future profits and consequently the performance effects of diversification depend on the rate of time preference of the market. The higher the market's rate of time preference, the worse is the performance effect of diversification at a time close to the moment of merging (at least in the example depicted in Figure 4 above).[66] The importance of the market's rate of time preference in the financial measure of performance can provide an explanation for Schoar's finding that diversified firms are more technically efficient (higher productivity) but are valued lower by the capital market. Concluding this issue then, giving a prominent role to the time dimension is an important direction for future empirical work on the relation between diversification and performance (cf. Grant, Jammine and Thomas).

2. A second persistent weakness in the research on the diversification-performance nexus is the measurement of diversification. The recent studies discussed above all use SIC-codes to determine either that a firm is diversified, i.e. being active in more than one SIC code, or the extent of firm diversification, i.e. the number of SIC codes a firm is active in. SIC codes however, reveal only little about technological distance, which as Jaffe (1986) shows plays such an important role in determining the usefulness of spillovers. Since spillovers are closely related to the of economies of scale and/or scope diversification could be meant to exploit, the concept of technological distance should be important in studies investigating the performance consequences of diversification. There obviously is a difference between a firm active in three or more four-digit SIC codes that are technologically close-by and a firm that is active in only two four-digit SIC codes that are

[66] The reverse situation would occur when diversification has short-term benefits but reduces profitability in the long run.

at great technological distance.[67] That is, a firm that is less diversified when SIC codes are used as the measure of diversification can have a position in technology space (technological position) (cf. Jaffe 1986) that is much more dispersed than the position of a firm active in more SIC codes. Similarly, a firm active in one four-digit SIC code may have a much more dispersed market position (cf. Jaffe 1988) than a firm active in two closely related four-digit SIC codes.[68] Thus, an ideal measure of diversification should be able to incorporate both technological position and market position, and depict them in a single number reflecting the extent to which a firm is diversified. For instance, using Jaffe's (1986) concept of technological distance to calculate market dispersion on the basis of the position in market space corrected for technological proximity of the different markets (see also Jaffe 1988). A firm that is active in three different but technologically related industries is then less diversified than a firm active in two completely different industries. Such a measure would have the further advantage that the extent of a diversifying move by the firm can easily be determined by comparing the - before and after - technology-corrected market position (the position in market space corrected for technological proximity of the different markets). In theory, a firm consisting of two technologically unrelated segments can enter a third industry with close technological ties to the two original industries and consequently decrease its dispersion as measured by its technology-corrected market position. In addition, such a measure would make it easier to determine the main motive underlying the decision to diversify. A diversification move, which brings on only little change in a firm's technology-corrected market position, is more likely to be an attempt to exploit economies of scope than an attempt to enter greener pastures.

Again, making use of a measure of diversification like a technology-corrected market position seems a fruitful direction for future empirical work on the relation between diversification and performance.

3. Other interesting directions of future research follow mainly from the productivity effects of diversification as investigated by Lichtenberg and Schoar. Similar to Adams and Jaffe, these studies examine the level of productivity using data from the Longitudinal Research Data (LRD). Possibly fruitful directions for future research would be to combine the two

[67] The technological position of a certain SIC code should be seen as the technological position of the average single segment firm active in the particular industry.

[68] Dispersion of the firm in technological and/or market space can, for instance, be calculated by a Herfindahl type measure created by simply taking the absolute value of the vector denoting a firm's place in technology and/or market space (cf. Jaffe 1986 and 1988).

types of studies, or more general, studies of the diversification-performance nexus and firm-level spillovers. For example, Adams and Jaffe have investigated whether diversification plays a role in their finding of R&D being diluted by the number of plants a firm comprises. The result was robust, but this test, as in fact Adam and Jaffe's entire study, was cross-sectional in nature. Schoar in contrast has found that the act of diversification leads to a decrease in productivity. An interesting question subsequently is if the performance consequences of diversification are effected by whether or not a diversifying action is accompanied by an increase in the number of plants the firm comprises. Similarly, throughout this chapter, and also in the studies of Schoar and Adams and Jaffe, age was found partly to determine productivity levels, and it thus seems that further investigating the effect of (firm/plant) age may render interesting results. One would like to know, for example, whether older firms are more diversified (as is likely) and whether older firms comprise more plants (as is likely).

3.5 Firm size and firm performance

The previous section has discussed the effect diversification has on a firm beyond that related to R&D and discussed in Section 3.3, namely the relation between diversification and firm performance. Just like diversification, size may influence a firm beyond the relations discussed in Section 3.3. As size contributes to the profitable exploitation of economies of scale and scope, resulting from, amongst others, knowledge's nonrivalrousness, a positive relation between firm size and firm performance (profitability, productivity, etc.) is expected. Similar to the previous section then, studying the effect firm size has on firm performance, seems a promising area of research for assessing the role knowledge, R&D and spillovers, play at the firm level. However, the same broad caveat as made in the previous section applies: the relation between firm size and firm performance is moderated by a range of variables and may consequently not be useful for actually drawing inferences regarding the role knowledge, R&D and spillovers, play in firms. In addition, studies that have investigated the relation between diversification and performance have by virtue of the close relation between diversification and firm size also implicitly investigated the relation between firm size and performance.

Therefore, in this section attention will be limited to some studies that have investigated the relation between firm size and the firm's level of productivity, e.g. Majumdar (1997) and

Idson and Oi (1999) (see also Lichtenberg discussed in the previous section, and Adams and Jaffe and Aw).

Majumdar (1997)

Majumdar investigates the relations between firm size and age on the one hand and performance on the other for a sample of 1020 Indian firms. He finds mixed results. Size has a negative performance effect when performance is measured as productivity and a positive effect when profitability is used as the measure of performance. In much the same fashion, age and productivity are positively related while age and profitability are negatively related. These results are found using several control variables such as diversity, which has no statistically significant effect, and marketing expenditures relative to total operational expenditures, which have a negative effect on productivity but a positive effect on profitability. Majumdar finds that these partly diverging results can be attributed to the institutional framework of the Indian economy, which demonstrates the importance of this type of framework in this and similar type of research.

Idson and Oi (1999)

Idson and Oi study wage and productivity differences between firms of different sizes. They find that workers in larger firms are both more productive and earn a higher wage than workers in smaller firms do. Workers in firms with between 1 and 49 employees are less than half as productive as workers in firms with more than 2,500 employees are. The wage of the workers in the smallest firms is slightly more than half that of the workers in the largest firms. However, when the actual number of hours worked is controlled for, their wage is slightly less than half that of their colleagues employed in the largest firms. A possible explanation is found in the finding that larger firms occupy different niches in the market as demonstrated by their early adoption of a wide range of new and advanced technologies.

The findings of the two studies discussed here and those of Lichtenberg and Aw discussed in earlier sections leads to one simple conclusion and that is that larger firms are more productive. An important caveat is in order, however, because firms may be larger exactly because they are more productive (cf. Nickell 1996). Interesting future work might be to investigate (further) the moderating effect of diverse variables like age, the adoption of advanced technology, and also to experiment with different measures of size (e.g. the relation between relative size in a product market and the level of productivity). Especially age is

expected to play an important role because: (i) it has a strong relation to firm size, (ii) younger firms are more likely candidates for failure (in the near future) than are older more established firms (see Section 3.1 and some of the references therein), (iii) learning by doing varies with firm age (cf. Bahk and Gort 1993), and (iv) technology adoption varies with age (and size) (cf. Dunne 1994 and Doms, Dunne and Roberts 1995). The simple finding itself is, however, already quite interesting because it is consistent with the level scale effect prediction of the models of growth without scale effects discussed in the previous chapter, e.g. Jones (1995b), Smulders and van de Klundert (1995), Young (1998), Li (2000), and Peretto and Smulders (2002). Thus, while at the aggregate level there is no relation between size and the (long-run) level of productivity (output per worker) such a relation does exist at the firm level. In Chapter 2, the reason why there was no level scale effect, i.e. the United States is not richer or more productive than Luxembourg, was said to be the (rapid) diffusion of knowledge (through spillovers) which limits potentially large differences in productivity between countries with similar capital and human capital endowments. Even then it was noted that diffusion is not perfect and certainly does not occur instantly, but that it is important nonetheless, amongst others, to explain the absence of level scale effects between countries of different size (see also footnote 14). Now level scale effects and diffusion seem to be important again.

3.6 Discussion and conclusions

This chapter (together with Appendices B, C and D) has discussed a broad collection of studies and their findings in order to gather information on the role knowledge, R&D and spillovers, play in firms. The rationale for studying these particular factors is that these are the factors deemed most important in the macro-level study of economic growth without scale effects; knowledge is created through R&D activities and the nonrivalrousness of knowledge is the driving force behind spillovers and ultimately growth scale effects. With more information on the role played by knowledge, R&D and spillovers, play at the firm level, ultimately the models of growth without scale effects may be tested.
Unfortunately, this idea of using the findings reported on in the previous sections of this chapter does not work out in practice. The findings do not lend themselves for such a demanding task simply because they are not hard and concrete enough. That is, almost every finding is accompanied by a large ‘but’ or a ‘however’. For instance, one of the main findings

is that firm-level spillovers matter, they are large and significant just like they are at the macro level (e.g. Frantzen). It is not clear, *however*, whether these spillovers are strong enough to lead to a scale effect in growth. There is some evidence that firm size leads to faster productivity growth (e.g. Pagano and Schivardi) but there is also evidence that spillovers are strongly diluted by the number of plants a firm comprises (e.g. Adams and Jaffe). Similarly, firm behavior and firm characteristics are related to R&D in a way that is consistent with the idea that knowledge through its nonrivalrous nature gives rise to economies of scale and scope (e.g. Montgomery and Hariharan 1991 and Cohen and Klepper 1996b). The performance effect of the exploitation of these economies of scale and scope, however, seems to be limited or even absent.

All this is not to say that nothing interesting and useful for the issue of scale effects in growth has come from the study of the various articles. In fact, quite the opposite is true as is evidenced by the following discussion, which starts with presenting the four most important conclusions that follow the work discussed - and remarks made - in this chapter and in which subsequently some additional concluding remarks are made.

1. There are various relations between firm characteristics that could/would form an interesting avenue for future research. That is, any relation between the following incomplete list of variables characterizing a firm may be an interesting subject for further study: R&D expenditures, R&D capital, R&D intensity, firm size, firm performance, productivity, productivity growth, number of firm segments, number of plants, number of plants per firm segment, firm age, plant age, segment age, etc. Most interesting for the issue of growth scale effects would be to examine further the relation between firm size and productivity growth (with the inclusion of moderating variables at both the firm and the industry level). As mentioned, the data needed for this type of study is readily available.
2. Life-cycle effects seem to be important. That is, (relative) age (of the firm or the plant) was shown to be related to firm size growth and productivity. In addition, looking for greener pastures was judged to be an important motive for diversification and in the search-theoretical framework of Matsusaka, it is able to account for the small and quite possibly negative performance effects of diversification. Hence, of the limited enumeration of relations between firm characteristics providing interesting avenues for future research given in the first concluding remark above, for the issue of growth scale

effects at every level of economic analysis (i.e. micro or macro), further study of the relations between age and other variables seems to be most promising.

3. Firms differ. Firms differ in size, their level of diversification, expenditures on R&D, productivity, etc. Hence, if anything, the idea of a representative firm used in all the models of growth discussed in this thesis is a vast simplification of economic reality.
4. There is evidence that points to the existence of some sort of dilution effect in which the value of knowledge measured by the extent of its productive use (i.e. R&D capital and R&D efforts) is limited through increases in the number of receivers the knowledge is spread over (cf. Adams and Jaffe) is a third interesting finding. This idea of dilution is very close to an effect moderating knowledge's nonrivalrousness and subsequently is destined to play a role in models that want to dispose of the prediction of a scale effect in growth of the early models of R&D-based growth (see for instance Peretto and Smulders).

Besides the fact that these four conclusions are important for the issue of growth scale effects, i.e. they may be crucial to the understanding of - and the ability to credibly model - growth without scale effects, they also provide food for further thought, especially the last one regarding the dilution of knowledge, and some additional remarks are required.

- A first of these additional remarks is that there is a lack of studies that have investigated (or can be used to investigate) the role knowledge, R&D and spillovers, play at the firm level using time-series data. This cross-sectional nature of most studies including that of Adams and Jaffe on the dilution of knowledge means that using their findings in the more dynamic context of scale effects in growth must be done carefully. Adams and Jaffe find that the level of productivity (of for instance a firm) depends not so much on the absolute amount of knowledge available to it, but rather on the relative amount available, i.e. the amount of knowledge per plant or something similar. However, the sort of scale effects that are of interest in this thesis are not level scale effects, rather the scale effects of interest are of a temporal nature; the rate of growth of a variable increasing with size (or not).[69] Empirical work that would shed the most light on this issue is not work on levels of productivity, but work along the lines of Jones (1995a), i.e. work in which country size changes, with changes (e.g. increases) measured using suitable measures like R&D expenditures, are related to the growth rate. Unfortunately, the study of Adams and Jaffe completely lacks a time component. Consequently, the question whether dilution plays a

[69] In fact, as mentioned in Chapter 2, both the one-R&D-sector and the two-R&D-sector models of growth without scale effects exhibit a level scale effect. That is, these models predict that larger economies in the steady state have larger income per capita and (labor) productivity than smaller economies.

role in limiting the scale effect in growth, i.e. the growth rate increasing with size (suitably measured; cf. Backus, Kehoe and Kehoe), is a question that needs further answering. There are more benefits of using time-series data however: (i) fixed effects that are specific to a plant, a firm, an industry etc., play a smaller role in regression analysis because they can be accounted for by using time-series data on economic subjects (e.g. firms) that are very similar without this use of a small cross-sectional sample leading to disadvantages related to limited amounts of data, and (ii) life-cycle effects are deemed important and these effects do not get proper attention in cross-sectional analysis.

- A second remark is that pure knowledge spillovers (can) have dynamic and/or static effects, i.e. pure knowledge spillovers (can) have a positive externality regarding the effectiveness of future R&D effort in increasing productivity (dynamic) and/or a positive externality regarding the (current) level of productivity (static). The previous remark regarding time-series data can be summarized as to say that time matters. This follows from the difference between level scale effects, larger economies (or firms, etc.) are more productive than smaller ones, and growth scale effects, productivity in larger economies (or firms, etc.) grows faster than in smaller ones. It was discussed in Chapter 2 that scale effects in growth, which are the type of scale effects that are of interest in this thesis, result from the nonrivalrous nature of knowledge giving rise to spillovers. The mechanism generating growth scale effects in the early models of R&D-based growth subsequently is that a larger economy uses more resources in the production of knowledge, and that the larger an economy's knowledge stock (accumulated past R&D efforts) is the more productive its current and future R&D efforts are. In this mechanism the spillovers are intertemporal and have dynamic effects. Knowledge and the spillovers associated with its nonrivalrousness may, however, very well have static effects as well. To see how this works, take the discussion of advertising and R&D as similar types of investments in footnote 44 above. It was argued there that for a firm these investments can have the same effect of increasing productivity - that is - of increasing the level of productivity. In addition, both advertising and R&D have spillovers (externalities). Expenditures on advertising by one firm spillover to other firms through the effect they have of increasing demand for a class of products (use toothpaste to prevent caries) and/or for complementary products (coffee and sugar). Because of this demand increase for the products of other firms, the prices of the products of these other firms increase ultimately raising their level of productivity. Similarly, expenditures on R&D by one firm create knowledge that raises its productivity but when spilling over to other firms also increases

the level of productivity of these other firms. However, it was also argued in footnote 44 that there is a difference between investing in advertising and investing in R&D, which is that it is improbable that past advertising expenditures (advertising capital) increase the effectiveness of future advertising expenditures. This touches exactly upon the issue of intertemporal spillovers and their dynamic effects and combined the above leads to the conclusion advanced here that (pure) knowledge spillovers (can) have two types of effects, static (level) effects and dynamic (growth) effects, i.e. pure knowledge spillovers (can) have a positive externality regarding the (current) level of productivity (static), and/or a positive externality regarding the effectiveness of future R&D effort in increasing productivity (dynamic). For the issue of scale effects in growth, this conclusion has some interesting implications. For instance, the dilution of knowledge as found by Adams and Jaffe seems to be more suited to limit static spillovers simply because their findings are based on a regression analysis of the level of productivity (rather than growth of productivity). This is not to say, however, that the concept of dilution does not play a role in limiting dynamic (intertemporal) spillovers, rather it again leads to the conclusion that more time-series analysis is required this time to study more specifically the dynamic effects of knowledge spillovers (at the firm level).

- A third (and final) remark, is that placing the dilution effect in a dynamic environment of demand for a product taking off and ultimately declining could provide a mechanism by which scale effects are absent in growth that is more firmly established in microeconomic facts and phenomena than the mechanisms used by the models of growth without scale effects discussed in the previous chapter. This remark follows from one of the ideas behind this chapter namely that looking at microeconomic regularities could somehow help assess the validity of proposed models of growth without scale effect, i.e. provide a tests. So far, the Peretto-Smulders model scores best in the test for it provides the most microeconomic foundation for its prediction of growth without scale effects through its incorporation of the dilution effect and relating it to technological distance. This is not to say, however, that Solow's (1994) critique does not, to some extent, apply to this model as well, for several issues, which touch upon the four conclusions presented above and the previous two remarks, arise. The Peretto-Smulders model does, however, provide a suitable framework to discuss these issues. In this model firms participate in networks, which can be viewed as equivalents of the technological clusters of Jaffe (1986). Within such a network/cluster firms use much the same knowledge, i.e. firms are technologically close, which in turn implies that dilution is limited. One subsequently expects that

spillovers within networks, as opposed to macro-level spillovers, are strong enough to lead to growth scale effects because they are not diluted through technological distance as macro-level spillovers are. That, nonetheless, is not the case and follows from the way Peretto and Smulders model firms' participation in the networks: rather than that a firm participates in a given network for prolonged periods (more than one), each firm randomly joins a different network each period (p. 612). This effectively eliminates the intertemporal R&D spillovers needed for growth to exhibit a scale effect because each firm is bound to join a network in which, due to the effect of technological distance, the knowledge it gathered in the previous period(s) has severely limited usefulness. At this point however, the Peretto-Smulders model takes a strong leap away from economic reality (cf. Solow 1994). It seems far more realistic that a firm's choice of which network to participate in is not determined at random but rather is 'endogenous', determined by economic factors. One would expect each firm to decide each period between two concrete options: (1) stay active in the same network and benefit from relatively useful knowledge accumulated in the past, or (2) shift its activity to another network in which its knowledge stock may be of limited use but which has other benefits. Given the amount of knowledge a firm may have accumulated in a certain network it will not change networks without a strong incentive to do so and what incentive could be stronger than that the current network is in decline and that hence other networks simply provide greener pastures. That is, there are obviously life-cycle effects playing a role in the decision of a firm of which network to be active in: the older a certain product market is or similarly the longer a firm is active in a certain product market, the less likely it is that this market is at the beginning of its life cycle and (the take-off phase) and the more likely it is that this market is at the end of its life cycle (the decline phase) (see also Chapter 6), and product markets that are at the beginning of their life cycle generally provide greener pastures than those at the end of their life cycle. In addition, this idea of firms changing their activities in response to declining and emerging opportunities (e.g. in response to a product market in the decline stage of its life cycle and a product market in the take-off stage of its life cycle) is supported by (i) one of the motives of diversification presented above (greener pastures) and the search-theoretic explanation of the diversification discount developed by

Matsusaka, and (ii) the fact that firms differ[70]. Most interesting however, is that just like random network participation, endogenous network participation may limit (intertemporal) spillovers to a sufficient extent to cause growth not to exhibit a scale effect. That is, the assumption of random network participation may not be crucial to the main result of the Peretto-Smulders model of growth without scale effects and the same result might be achieved, i.e. intertemporal spillovers may still be limited sufficiently, when network participation is endogenous. Placing the dilution effect in a dynamic setting of firms continuously changing there lines of business, for instance in response to threats and opportunities associated with life cycle stages, could achieve the same feat of effectively eliminating the intertemporal R&D spillovers needed for growth to exhibit a scale effect, without the favorable (unearned) assumption (cf. Solow 1994) of random network participation being required. Rather, life-cycle effects provide the foundation for the idea that firms indeed do continuously change there lines of business so that incorporation of these effects not only makes the models more realistic, i.e. complying with more empirical (micro-level) regularities, but actually is an important building block of the mechanism by which growth does not exhibit a scale effect proposed here. Hence, endogenizing network participation in the Peretto-Smulders model would constitute a mechanism by which growth does not exhibit a scale effect that has a stronger microeconomic foundation than any of the mechanisms driving the main result of growth without scale effects of the models discussed in the last section of the previous chapter. In addition, making a model consistent with observed empirical regularities not only strengthens its credibility, allowing it to pass more tests (cf. Thompson 2001b), but these regularities may actually provide important building blocks for modelling other observed facts and phenomena.[71]

With the four conclusions and these final remarks in mind the most interesting conclusion of this chapter is then that technological distance, dilution and life-cycle effects are important in the issue of growth scale effects and they return in the next chapter.

[70] Endogenous network participation, i.e. life-cycle effects, impl(ies)(y) that each firm is different as evidenced by a different history of network participation and a different stock of knowledge accumulated in the past. This in turn introduces path dependency, which does not go well with the representative agent approach to modelling R&D-based growth taken by for instance Peretto and Smulders. A different, less formal and more appreciative modelling approach is consequently required (cf. Nelson and Winter 1982).

[71] Note, however, that the only model that can explain all observed empirical regularities is the world itself.

4 Growth without scale effects and structural change

The fact that scale effects in growth are absent at the macro level despite the fact that knowledge is at least partly nonrivalrous, which is not only a logical assumption but is also evidenced by the findings of some of the studies discussed in the previous chapter, forms quite a puzzle.

In the first section (4.1) of this chapter a simple (toy) model that has a lot in common with the models of growth without scale effects discussed in Chapter 2 will be developed. The difference with these models lies not only in its simplicity[72], but also in the incorporation of phenomena deemed important from the study of knowledge, R&D and spillovers, at the firm level of the previous chapter, e.g. knowledge dilution and life-cycle effects. In Section 4.2 subsequently, theoretical literature closely related to the model and its main assumptions are discussed while similarly in Section 4.3 empirical work is discussed. Section 4.4 ends the chapter and provides a discussion and some conclusions.

4.1 A simple model

In their cross-sectional analysis of the scale effects issue, Backus, Kehoe and Kehoe find no evidence on scale effects in growth when looking at aggregate production (GDP), but do find evidence on scale effects when looking at a single sector, manufacturing. That is, they find that larger manufacturing sectors grow faster than smaller manufacturing sectors do, but that larger economies do not grow faster than smaller economies do. These seemingly contradictory findings (can) have an intuitive explanation, which is that the manufacturing sector forms so small a part of the total economy that its contribution to a growth scale effect at the macro level gets lost in the aggregate.[73] This is not to say, however, that growth in the other sectors of the economy do not exhibit growth with scale effects as well. It just means that the growth rate of the economy at large, calculated as the weighted growth rates of the various sectors that constitute the economy, does not exhibit a scale effect. That all sectors of the economy can exhibit growth without scale effects, while aggregate growth shows no evidence of a scale effects seems even more contradictory than the finding of Backus, Kehoe

[72] In fact, the model is so simple that the term model must be seen as an exaggeration and a better term would be 'mathematical stylization'.
[73] The growth of GDP is of course equal to the weighted growth of the various sectors of the economy.

and Kehoe who find evidence for a scale effect in growth only for the manufacturing sector. In this section, however, a simple stylized model is developed, which is capable of reconciling these apparent contradictions.

The idea is that demand (and/or output) shift from goods produced in one sector to goods produced in another sector, i.e. there is structural change. With structural change taking place, a sector can exhibit growth scale effects but because demand shifts to a (new) less productive sector relatively fast, the scale effects simply vanish at the aggregate level. As the model is very simple, the - obviously complex - interaction among the variables underlying this mechanism of structural change is not explicitly modelled. Instead, to keep the intuitive nature of the mechanism just described, the variables that are of interest are straightforwardly defined as a function of time in a way that is exogenously given. More concrete, the whole demand side of the economy described by the model, i.e. the maximizing behavior in which consumer preferences but also (and especially) factors like income, (relative) prices, productivity growth (in turn mutually related) etc. play a large role, is captured in a simple exogenously given time-pattern of demand. In addition, the process by which technological progress occurs (e.g. through R&D) is not modelled but rather sectors are simply assumed to have growth rates that increase over time, i.e. sectors are assumed to exhibit a scale effect in growth. The model is subsequently analyzed on the basis of numerical examples.

4.1.1 The basic framework

Assume an economy that at time t = 0 has one sector (sector 1) that produces a single good (good 1). Over time, the economy gets better at producing this good.

The rate of technological progress in sector 1 (g_1) increases over time according to:

$$g_1(t) = \max\left\{0, g_{0,1}[1-e^{-\tau_1 t}]\right\} = g_{0,1}[1-e^{-\tau_1 t}] \qquad (4.1)$$

where $g_{0,1}$ is the (exogenously given) maximum rate of technological progress of sector 1, and $\tau_1 > 0$ is a parameter that determines the speed at which the rate of technological progress (TFP growth) of sector 1 approaches its maximum ($g_{0,1}$). Technological progress is modelled in this way for all sectors and thus, assuming that the sectors grow in size (suitably measured) over time, growth of TFP in each sector exhibits some sort of scale effect, i.e. increases over time as size (again suitably measured) increases (though at a decreasing rate) (see Figure 5 below).

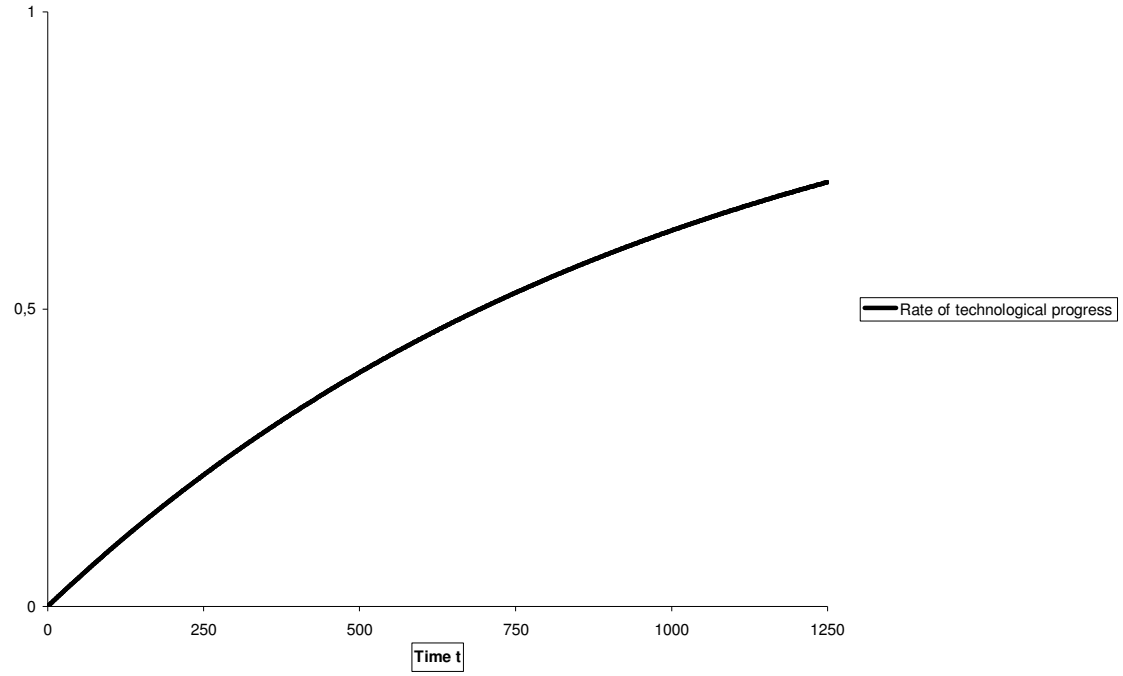


Figure 5, The growth rate of sectoral TFP as a function of time.[74]

Using the rate of technological progress, total factor productivity in sector 1 (TFP_1) can be written as:

$$TFP_1(t) = TFP_{0,1}e^{g_1 t} \tag{4.2}$$

where $TFP_{0,1}$ is total factor productivity of sector 1 at time t = 0 (an exogenously given constant).

After some time a new good (good 2) is invented and at time t = T, demand conditions are such that production of this new good becomes profitable. The consequent start of production of this new good creates a second sector in the economy (sector 2). The rate of technological progress in the second sector (g_2) follows a similar path as that in the first sector and the same goes for total factor productivity in this sector (TFP_2):

$$g_2(t) = \max\{0, g_{0,2}[1 - e^{-\tau_2(t-T)}]\} \tag{4.3}$$

$$TFP_2(t) = TFP_{0,2}e^{g_2 t} \tag{4.4}$$

where the variable T, as discussed above, is an exogenously given time lag. The introduction of new goods (or sectors) is very important for the results of this simple model and in this respect it does not differ much from the two-R&D-sector models of growth without scale

[74] For this figure the parameter values have been set to: $g_{0,1} = 1$, $\tau_1 = 0.001$, and $TFP_{0,1} = 1$.

effects discussed in Chapter 2 in which the introduction of new goods is crucial in eliminating growth scale effects as well.

Total production in a sector is determined by sectoral total factor productivity and the amount of resources that each sector has available. There is assumed to be only one kind of productive resource that is available in a constant amount set equal to 1. This resource is distributed over the production of good 1 and good 2 due to the interaction between factors that are internal to the economy; there is no trade. In order to keep the model simple, it is assumed that these factors, consumer preferences, income, (relative) prices, productivity growth and so forth, work together in such a way that the distribution of the productive resource follows a simple pattern over time. That is, as mentioned above, the demand side of the economy is not explicitly modelled and instead the factors important in determining demand are simply assumed to be interacting in such a way that demand and consequently the distribution of productive resources evolves over time according to:

$$\sigma_{12}(t) = \min\left\{1,[1-e^{-\tau_{12}(t-T)}]\right\} \tag{4.5}$$

where $0 \leq \sigma_{12} \leq 1$ is the percentage of the productive resource used in the production of good 1, and $\tau_{12} > 0$ is a parameter that determines the pace at which demand and thus the distribution of the productive resource shifts from sector 1 to sector 2. The shift of demand and the corresponding structural change does not occur until good 2 becomes available at time t = T. Thus, rather than that good 2 is invented at time t = T, good 2 can also be seen to have been available from the start (t = 0), but demand for it does not reach a positive level until time t = T, for instance because only then income has reached a high enough level. This also means that the process by which productivity in this sector increases does not start until time t = T, i.e. firms invest in R&D for the new product only when this new product actually is demanded. In this respect, technological progress in this model very much seems to be the result of learning by doing. Learning by doing, however, does not lead to technology progressing at an increasing rate (cf. Bahk and Gort) and hence technological progress in this model is most closely associated with R&D, which in turn is associated with intertemporal knowledge spillovers resulting from knowledge's nonrivalrousness.

From equation (4.1) through (4.5), two important assumptions of the model are clear:

1. there are demand shifts (i.e. structural change) and at time t = T, sector 2 'steals' away output (demand) from sector 1,
2. the level of productivity in the new sector is lower than in the old sector.

These features form the driving force behind its result that growth scale effects are absent at the aggregate level while they do exist at the sectoral level. The idea is namely that demand shifting from a sector with a relatively high level of productivity to a sector with a relatively low level of productivity sector can cause growth scale effects to get lost in the aggregate level, at least for some time. Of course demand shifts do not occur exogenously as assumed here but as the result of maximizing behavior of the consumers in which factors like consumer preferences and income, (relative) prices, productivity growth (in turn mutually related) etc. play a large role. The assumption of demand shifting from one sector to another must thus be seen as a summary of assumptions regarding consumer behavior the result of which is simply that the composition of total demand (or consumption or output) changes over time.[75]

To complete the model, total demand for goods produced in sector 1 (Y_1) and sector 2 (Y_2) respectively equals total production in sector 1 and sector 2 and can be written as:

$$Y_1(t) = \sigma_{12} TFP_1 \tag{4.6}$$

$$Y_2(t) = (1-\sigma_{12}) TFP_2 \tag{4.7}$$

and total production (Y) and demand is given by:[76]

$$Y(t) = Y_1 + Y_2 \tag{4.8}$$

That the model developed in equation (4.1) through (4.8) can be used to show that growth at the aggregate level does not have to exhibit a scale effect even though sectoral growth does exhibit such an effect can be clarified by a numerical example. The results of this example are depicted in Figure 6, which graphs the natural logarithm of TFP_1, TFP_2, and Y, as a function of time t.

[75] Note that, so far, the model developed here does not differ much from the models of R&D-based growth (with or without scale effects) discussed in Chapter 2. There too new goods are introduced that replace older ones. The main difference is that in the present model demand for older products remains, at least for some time, despite the introduction of newer products. In the models discussed in Chapter 2 there is only one final goods sector and not several existing alongside one another. Consequently, sectoral differences in the level - or growth - of productivity are not existent in these models either.

[76] As population growth is non-existent in this model, total output (GDP) is directly proportional to per-capita GDP. Similarly, because the amount of resources available is constant, growth of total output denotes growth of (total factor) productivity.

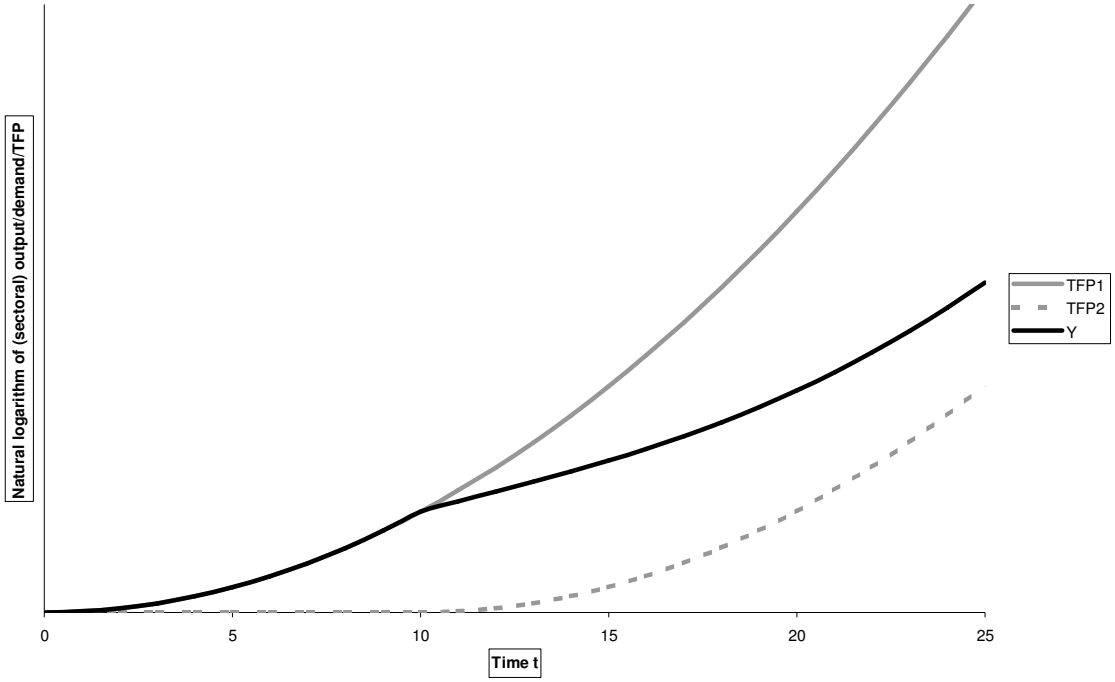


Figure 6, A numerical example of a simple two-sector-model of growth with sectoral demand shifts.[77]

From Figure 6, it is immediately clear that the sudden demand shift and concomitant redistribution of productive resources from sector 1 to the less productive sector 2 at time t = 10 affects the growth rate of total output (Y). It is not until approximately time t = 20 that the growth rate of per-capita income has overcome this effect. Between time t = 0 and t = 20 growth in per-capita GDP follows an almost linear path in which no scale effects are apparent, even though growth scale effects do exist at the sectoral level.[78] In addition, this path exists despite the fact that the two sectors are identical except for the time at which production (and growth of total factor productivity) starts. That growth in this simple toy model does not exhibit a scale effect is confirmed in Figure 7 below, which depicts the rate of growth of total output for the model for the period t = 0 till t = 25.

[77] For this numerical example the values of the constants have been set to: $g_{0,1} = g_{0,2} = 1$, $\tau_1 = \tau_2 = 0.001$, $TFP_{0,1} = TFP_{0,2} = 1$, T = 10, and $\tau_{12} = 0.1$.

[78] Between time t = 0 and time t = 20, the linear regression estimate of the relation between the natural logarithm of Y and t is: $\ln Y = 0.0116t - 0.0218$ (R^2 adj. = 0.9870).

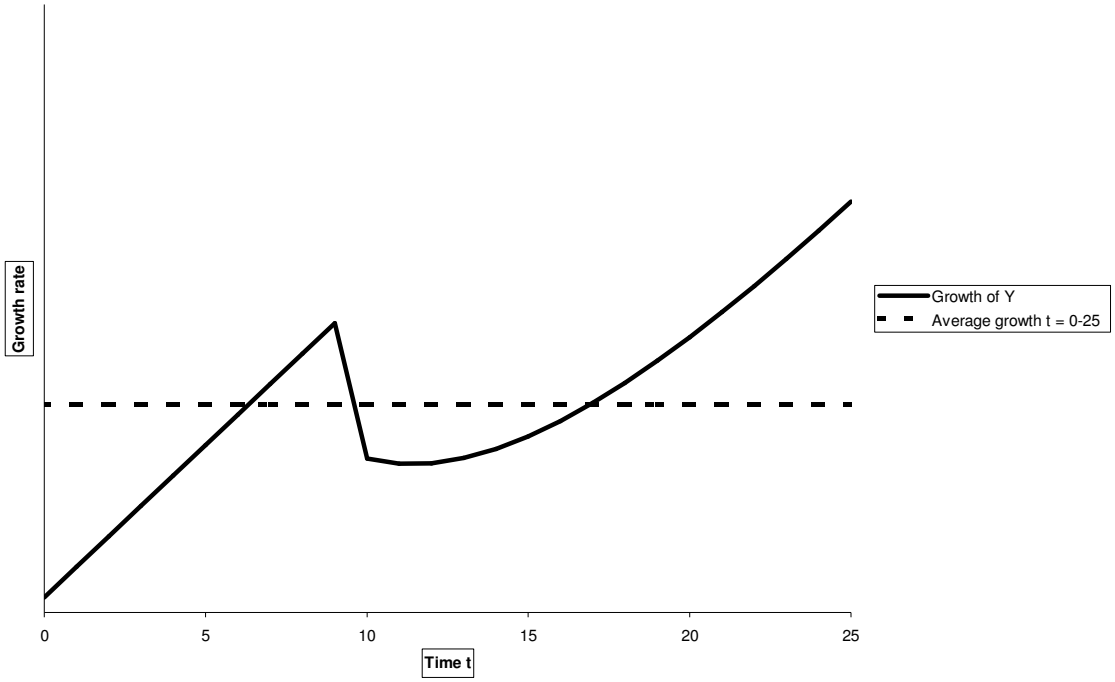


Figure 7, Growth rate of total output in the simple two-sector-model.[79]

As Figure 7 shows, growth has some sort of cyclical movement around its mean, and only after a prolonged period of time ($t > \pm 20$) does growth have a tendency to increase over time. Hence, the conclusion so far is that structural changes (in turn the result of maximizing behavior of the consumers), can, at least temporarily, lead to growth without scale effects. This result depends not only on the actual shift of demand, however, but also on the fact that it shifts from a sector with a relatively high level of productivity to a new sector that has a relatively low level of productivity.

In addition, the model has another feature that intuitively seems important: the speed at which demand shifts towards the new sector. Indeed, demand has to shift to a sector with a lower level of productivity at the right speed to cause both growth - and the level - of productivity at the aggregate level to be constrained. The feature that growth in the new sector is lower than in the old sector, which also seems important intuitively, on the other hand is not particularly important. As long as the productivity disadvantage of the new sector continues (despite this new sector's catch-up rate of growth) both growth - and the level - of productivity at the aggregate level are constrained. That is, the effect of a fast rate of productivity growth in the new sector can be (is) counteracted by an even faster shift of demand to this still less

[79] The parameter values used are the same as for the previous figure, Figure 6 (see footnote 77). Further note that using size, i.e. total output, as independent variable (instead of time) does not change the basic path of the growth rate.

productive sector. That this mechanism actually works is probably best shown in Figure 7 from time t = 10 to time t = 13. During this time, the rate of growth is roughly constant indicating that the several forces that determine growth of aggregate level productivity are nearly balanced.

4.1.2 A three-sector-model

This paragraph will extend the model of the previous paragraph to incorporate a third sector. The goal of the three-sector model is to show that introducing more sectors can lead to growth without scale effects for prolonged periods of time. The model takes as a starting-point equation (4.1) through (4.5).

After some time a new good (good 3) is invented and at time t = 2T, production of this new good starts, thus creating a third sector in the economy (sector 3). The rate of technological progress in the third sector follows a similar path as that in the other two sectors and the same goes for total factor productivity:

$$g_3(t) = \max\left\{0, g_{0,3}[1 - e^{-\tau_3(t-2T)}]\right\} \tag{4.9}$$

$$TFP_3(t) = TFP_{0,3}e^{g_3 t} \tag{4.10}$$

As in the previous model, total production in a sector is determined by sectoral total factor productivity and the amount of the productive resource that each sector has available.

Again, the complex interaction between income, productivity growth, (relative prices) and so forth, is not modelled and instead the distribution of the productive resource over the sectors is simply stated as a function of time:

$$\sigma_{23}(t) = \min\left\{1, [1 - e^{-\tau_{23}(t-2T)}]\right\} \tag{4.12}$$

where $0 \leq \sigma_{23} \leq 1$ is the percentage of productive resources used in the production of good 1 and 2. Thus again, rather than that good 3 is invented at time t = 2T, good 3 can also be seen to have been available from the start (t = 0) but demand for it does not reach a positive level until time t = 2T, for instance because then income has reached a high enough level.

Subsequently, the total production - and total demand for goods produced - in each sector can be written as:

$$Y_1(t) = \sigma_{23}\sigma_{12}TFP_1 \tag{4.13}$$

$$Y_2(t) = \sigma_{23}(1 - \sigma_{12})TFP_2 \tag{4.14}$$

$$Y_3(t) = (1-\sigma_{23})TFP_3 \tag{4.15}$$

Finally, total demand and production is given by:

$$Y(t) = Y_1 + Y_2 + Y_3 \tag{4.16}$$

Like the model developed in Section 4.1, the model developed in these equations and in equation (4.1) through (4.5) leads to growth that at least for some time does not exhibit a scale effect at the aggregate level, while sectoral growth does. Similar to Figure 6, Figure 8, graphs the natural logarithm of TFP_1, TFP_2, TFP_3, and Y, as a function of time in a numerical example, to show this property of the model.

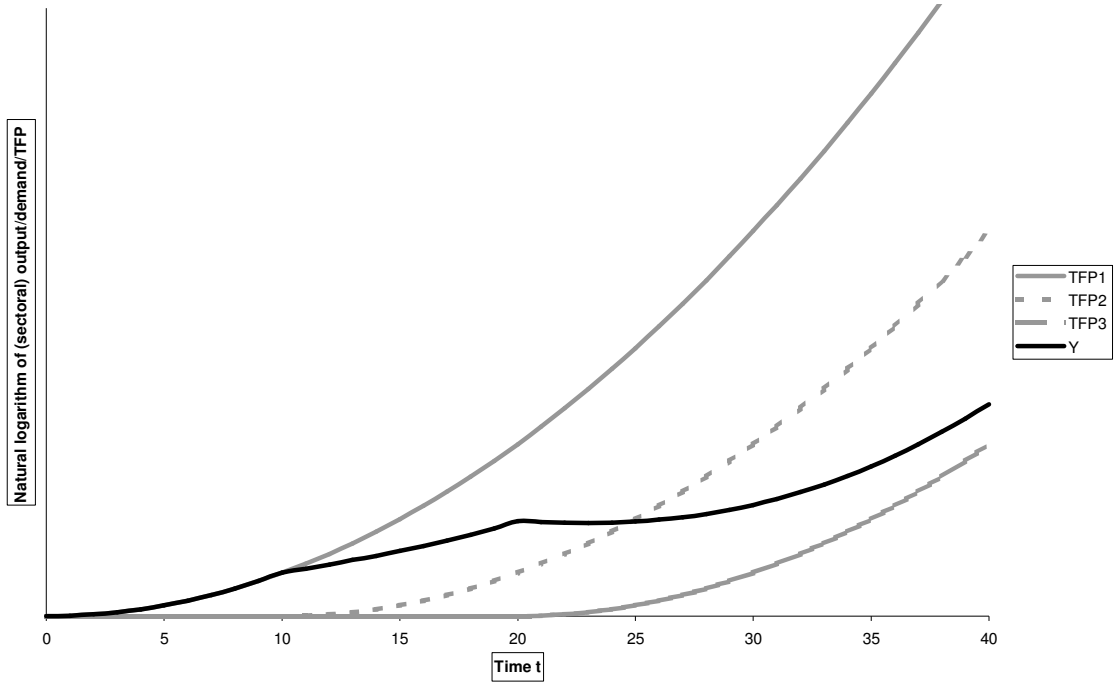


Figure 8, A numerical example of a simple three-sector-model of growth with sectoral demand shifts.[80]

Figure 8 clearly shows that the growth rate of total output (Y) is influenced by the redistribution of productive resources following the shift in demand from goods produced in one sector to goods produced in another sector (structural change). This changing composition of consumption (output) causes the growth rate of GDP to be nearly constant between time t = 0 and time t = 35, due to the demand shifts starting at time t = 10 and time t

[80] Again, the sectors are assumed to be identical. The values of the constants have been set to: $g_{0,1} = g_{0,2} = g_{0,3} = 1$, $\tau_1 = \tau_2 = \tau_3 = 0.001$, $TFP_{0,1} = TFP_{0,2} = TFP_{0,3} = 1$, T = 10, and $\tau_{12} = \tau_{23} = 0.1$.

= 20 respectively.[81] This is confirmed in Figure 9 below, which depicts the rate of growth of total output for the model.

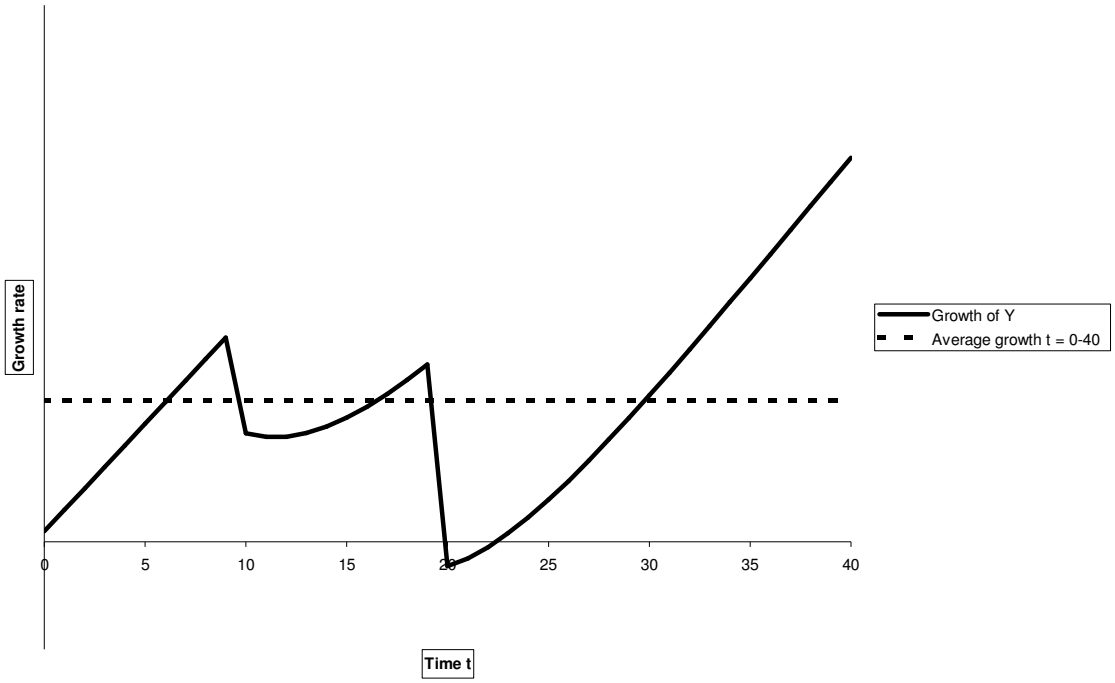


Figure 9, Growth rate of total output in the simple three-sector-model.[82]

Compared with the behavior of the growth rate in the two-sector model graphed in Figure 7, the period over which no scale effects can be found in aggregate production is lengthened in this model with an extra sector (three in total). Hence, the conclusion that prolonged structural changes caused by maximizing behavior of consumers in which factors like consumer preferences and income, (relative) prices, productivity growth etc. play a large role, can lead to growth without scale effects seems justified. Note, however, that now not only demand shifts, differing productivity levels and the speed of demand shifts, are important. The relative timing of the actual creation of the new sectors is important as well. If for instance demand for product 3 does not start at time T = 20 but at time T = 30, the picture changes drastically and it is no longer so obvious that growth at the aggregate level does not exhibit a scale effect.[83] Finally, Figure 9 shows a cyclical movement of growth around its (or a) mean in which growth for a brief period is actually negative. This cyclical behavior is not only

[81] Between time t = 0 and time t = 35, the linear regression estimate of the relation between the natural logarithm of Y and t is: $\ln Y = 0.0094t - 0.0039$ (R^2 adj. = 0.9717).
[82] The parameter values used are the same as for the previous figure, Figure 8 (see footnote 80).
[83] Of course, the time of introduction of the new product was also important in the two-sector model of the previous paragraph. However, it is more important in the three-sector model because this model's aim is to show that continued introduction of new products can be the reason why ultimately growth at the aggregate level does not exhibit a scale effect while at the sectoral level it does.

consistent with Figure 7, but also - and more so than Figure 7 - with actually observed behavior (see for instance Figure 1).

4.1.3 The three-sector-model with spillovers

One reason why the models developed previously do not exhibit scale effects is probably that knowledge accumulated in one sector, e.g. sector 1, has no influence on productivity in a next sector, e.g. sector 2. It was simply assumed that knowledge used in the production of one good is not useful in the production of another good, which must be seen as an extreme form of dilution of knowledge through technological distance (see Chapter 3).
The model developed in this paragraph does take the effect of spillovers from one sector to another (asymmetrically) into account. TFP in a sector is now modelled not only as a function of technological progress (g), but also as a function of TFP in the preceding sector. The model is again a three-sector model and equation (4.4) and (4.10) are now rewritten to:

$$TFP_2(t) = TFP_1^{\phi_{12}} e^{g_2 t} \tag{4.17}$$

$$TFP_3(t) = TFP_2^{\phi_{23}} e^{g_3 t} \tag{4.18}$$

where $\phi_{12} > 0$ and $\phi_{23} > 0$ are parameters that determine the extent of the spillover. Figure 10 below graphs the natural logarithm of TFP_1, TFP_2, TFP_3, and Y, as a function of time.

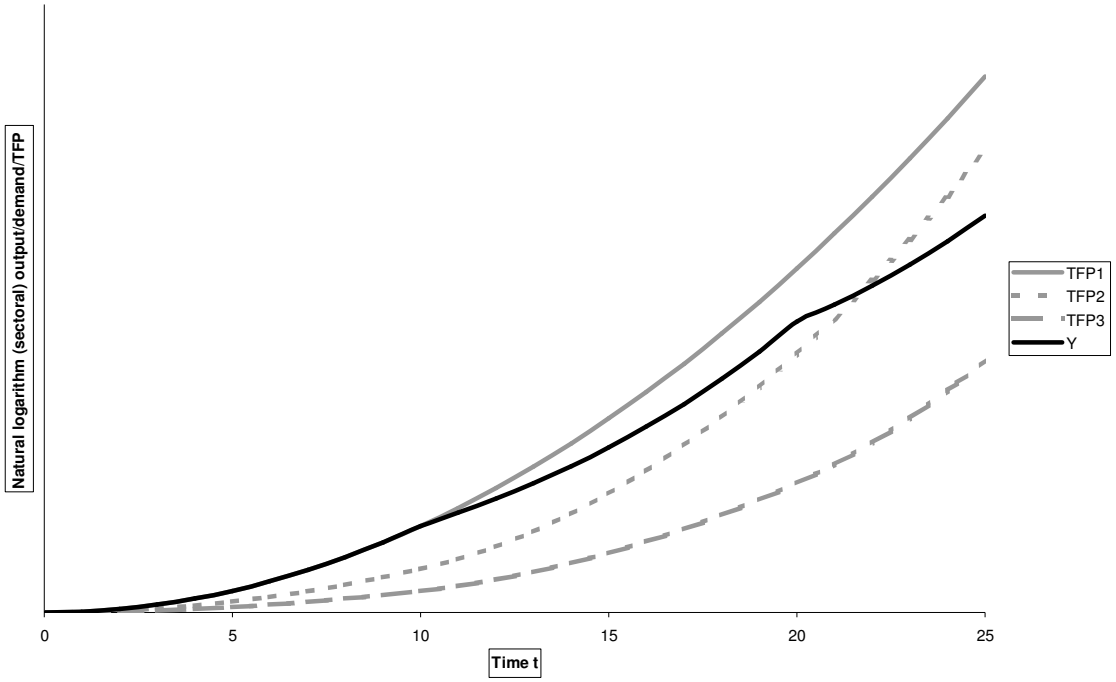


Figure 10, A numerical example of the simple three-sector-model with spillovers.[84]

The models depicted in Figure 8 and Figure 10 respectively only differ in the extent of intersectoral (level) spillovers. Compared with the three sector model graphed in Figure 8, the period over which no scale effects can be found in aggregate production in Figure 10 is severely reduced and overall the conclusion must even be that growth does exhibit a scale effect.[85] Hence, the conclusion is that structural changes can lead to growth without scale effects at the aggregate level, but that a mechanism that limits spillovers, i.e. dilution through technological distance, is important for obtaining this result. The larger the intersectoral spillovers are (i.e. the shorter technological distance between the respective sectors is), the larger is also the starting level of productivity in the newly created sector, and the less likely it is that structural change (changing composition of demand) causes scale effects to vanish at the aggregate level. That the extent of intersectoral spillovers indeed is somehow limited and that consequently the level of productivity in the new sector is lower than in the already existing sector(s) is, however, not at all unlikely. This is even more so when the concept of dilution (through technological distance) is extended to incorporate an idea like product-specific learning by doing (compare firm-specific knowledge; Smulders and van de Klundert).

[84] Again, the sectors are assumed to be identical. The values of the constants have been set to: $g_{0,1} = g_{0,2} = g_{0,3} = 1$, $\tau_1 = \tau_2 = \tau_3 = 0.001$, $TFP_{0,1} = TFP_{0,2} = TFP_{0,3} = 1$, $T = 10$, $\tau_{12} = \tau_{23} = 0.1$, and $\phi_{12} = \phi_{23} = 0.5$.

[85] Nonetheless, a linear model fits the behavior of output (Y) as a function of time rather well (at least for short periods). Between time t = 0 and time t = 30, the linear regression estimate of the relation between the natural logarithm of Y and t is: $\ln Y = 0.0218t - 0.0895$ (R^2 adj. = 0.959).

Besides as the result of R&D, productivity may increase through learning by doing (i.e. the accumulation of tacit knowledge) which by its very nature is closely related to a single product and cannot be expected to be very useful in the production of another product without some adaptation in the form of doing and consequently learning being required. Whether or not the broad dilution effect creates enough of a productivity difference between sectors to lead to growth without scale effects subsequently depends on other features of the model, namely the speed at which demand shifts between sectors and the relative timing of the introduction of new products (sectors). The idea forwarded in this section is of course that these factors, i.e. demand shifts and the speed at which these shifts occur (in turn the result of consumer behavior and the factors that determine this behavior), the introduction of new products and the broad dilution effect, work together to cause growth scale effect to be absent at the aggregate level while such effects might at exist at a sectoral level.[86]

A final remark regarding the three-sector model with spillovers concerns the nature of the sectoral spillovers. As mentioned in the previous chapter, spillovers (can) have dynamic and/or static effects. The spillovers implemented in the model here differ from the ones discussed in Chapter 2 and do not have dynamic intertemporal (growth) effects but static (level) effects. That is, the spillovers in the model and discussed here have a positive externality regarding the level of productivity but are not expected to increase the effectiveness of future R&D effort. Consistent with the evidence presented in the previous chapter (i.e. Adams and Jaffe) the static spillovers in the model are diluted: knowledge only reaches its full productive potential when used in the production of the products for which it was originally conceived, and through technological distance sort of 'loses' part of this potential when used in the production of other products (see also Peretto and Smulders). This is not to say, however, that there are no spillovers with intertemporal effects in the model at all. Though not explicitly modelled, such spillovers do actually exist within a single sector where they cause sectoral growth to exhibit a scale effect, i.e. to increase over time as size (suitably measured) increases. One is subsequently inclined to think that intertemporal spillovers are limited in this model by confining them to a sector, whereas these spillovers are

[86] The introduction of new products and the level of productivity are of course also somehow related to consumer behavior. New products are only introduced if firms find it profitable to do so and the same goes for R&D activities aimed for instance at increasing the level of productivity. This touches upon the question of whether there is a demand-pull or a technology-push effect with the question undoubtedly being that both effects exist and that both effects are important (see for instance Schmookler 1966, Kamien and Schwartz 1982, Walsh 1984, and Kleinknecht and Verspagen 1990). Further note that income and relative prices are in turn related to productivity growth and consequently to R&D. In fact, all the factors important in the model, e.g. consumer preferences, income, relative prices, technological progress (R&D) etc., are all intricately related and combined they make for the absence of scale effects in growth.

logically expected to work across sectors just as well. This indeed is the case but because, as noted above, the growth rate of productivity in the new sector is not particularly important for the result of the model, the explicit modelling of their effect across sectors is omitted for reasons of clarity.

4.1.4 Concluding remarks

The simple toy model or mathematical stylization developed in the preceding three paragraphs has several important features driving its main result of absence of a scale effect in growth at the aggregate level while such effects may be present at the sectoral level. The two most important are that,

1. there is structural change (i.e. demand shifts); when a new product (and sector) is introduced it steals away demand from the older sectors,
2. the level of productivity in the new sector is lower than in the older sectors for instance because of dilution of intersectoral level spillovers.

The idea of the model is simply that these and other features like the speed at which demand shifts from older products to the new product work together in such a way that growth at the aggregate level does not exhibit a scale effect while sectoral growth does exhibit such an effect. What is more, the idea is that all the factors important in the model, e.g. consumer preferences, income, relative prices, R&D (technological progress) and so forth, are highly intertwined and work together to establish this result (and that this is what actually happens in reality causing growth scale effects to vanish at the aggregate level).

Further support for the idea of an intricate relation (see footnote 86) between consumer preferences, income, relative prices, R&D (technological progress) and so forth, causing the absence of aggregate level growth scale effects comes from the behavior of demand (output) for the different products (sectors) over time. This behavior depicted in Figure 11 below follows a pattern that can best be described as a life-cycle pattern: demand goes to several (distinct) phases for instance three, (1) the take-off phase in which demand for the product takes off and grows almost exponentially (e.g. the take-off phase), (2) the maturity stage (or something similar) in which total demand reaches its maximum, and (3) the decline phase in which demand declines asymptotically to its final value, i.e. zero.

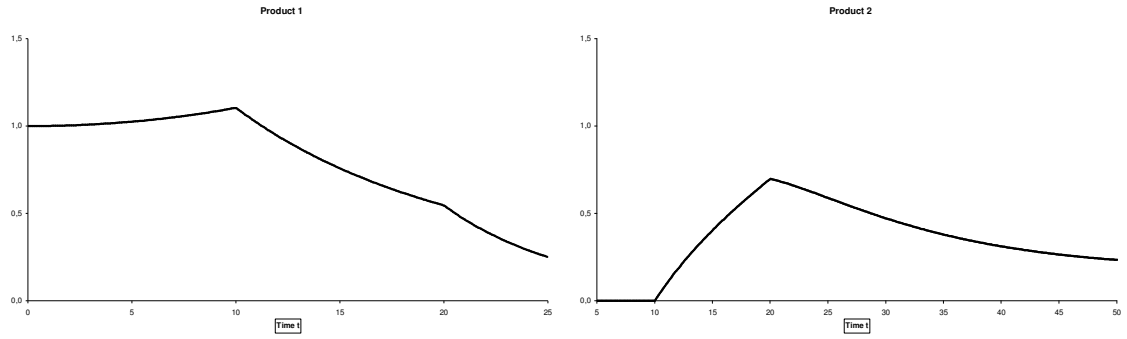


Figure 11, Demand/output for products 1 and 2 over time in the three-sector-model.[87]

The fact that products in the mathematically stylized toy model developed in this section go through a life cycle relates back to one of the conclusions of the previous chapter that life-cycle effects are important. What is more, the model is related to the idea of endogenous network participation developed in the previous chapter as a way to improve the Peretto-Smulders model. That is, the model developed here would allow for continued participation in a single network without intertemporal knowledge spillovers leading to aggregate-level growth scale effects in much the same manner as the model allows for sectoral scale effects while such effects are absent at the aggregate level. Just like demand for products from a certain sector needs to be limited, the duration of participation needs to be limited at some finite time. The similarity is subsequently that both demand and the duration of participation are constrained due to life-cycle effects, i.e. due to the life cycle (demand for) each product goes through. What this section has shown is thus that network participation does not need to be random to somehow limit intertemporal spillovers, but that the same result may be achieved, as argued in the previous chapter, when network participation is endogenous.
The driving force behind this result and ultimately causing the absence of aggregate level growth scale effects is of course the fact that products go through a life cycle with demand taking off and ultimately declining to its final value (e.g. zero), which in turn must be viewed as the result of the intricate relation between consumer preferences, income, relative prices, R&D (technological progress), etc.

This chapter continues with a discussion of theoretical and empirical work related to the above-given depiction of reality in sections 4.2 and 4.3 respectively.

[87] Again, the sectors are assumed to be identical. The values of the constants have been set to: $g_{0,1} = g_{0,2} = g_{0,3} = 1$, $\tau_1 = \tau_2 = \tau_3 = 0.001$, $TFP_{0,1} = TFP_{0,2} = TFP_{0,3} = 1$, $T = 10$, and $\tau_{12} = \tau_{23} = 0.1$.

4.2 Growth, income, and demand

The simple toy model or mathematical stylization developed in the preceding section has made important (favorable; cf. Solow 1994) assumptions, regarding (i) demand shifting from goods produced in one sector to goods produced in another sector (and the existence of an intricate relation between consumer preferences, income, relative prices, technological progress, etc.) especially, and regarding (ii) productivity differences between sectors to a lesser extent. There is, however, theoretical work that has discussed these admittedly strong assumptions (though seldom their combination) so they are not all that uncommon in literature. This section will discuss some of this work, namely Baumol (1967), Pasinetti (1981), Laitner (2000) and Aoki and Yoshikawa (2002), to provide some credibility to these seemingly far-fetched assumptions of the model.

Baumol (1967)

Baumol develops a framework in which the economy consists of two sectors. One sector has rising (labor) productivity (the technologically progressive sector), while in the other sector (labor) productivity is constant (the technologically stagnant sector). He subsequently argues that manufactured goods are more prone to being part of the technologically progressive sector, while services in which labor often is an end itself (e.g. a play) are more prone to being part of the technologically stagnant sector. The result is that, as Baumol states, "… manufactures are likely to continue to decline in relative cost and, unless the income elasticity of demand for manufactured goods is very large, they may absorb an ever smaller proportion of the labor force, which, if it transpires, may make it more difficult for our economy to maintain its overall rate of output growth." (p. 421). The difficulty of maintaining growth while services increase in importance as measured by their share in total output is subsequently often referred to as 'Baumol's disease' and his model is often called a model of unbalanced growth.

This remark by Baumol provides a strong link to the model developed in the previous section for it not only states that there may be differences in sectoral productivity growth rates and that the overall rate of output can be constrained, but more importantly it states that whether or not this actually occurs depends on the income elasticity of demand for goods produced in the technologically progressive sector. That is, consumer preferences and ultimately consumer behavior play a role in constraining (productivity) growth.

Pasinetti (1981)

Engel's law which states that "the share of food in total expenditures is inversely related to the household's income (or some other measure of its total resources)." (Houthakker 1987b, p. 143) is one of the most robust findings in economics (ibid.) and in fact demand has been found to follow Engel-type curves for a broad range of products (Deaton and Muellbauer 1983). Pasinetti is one of the firsts to recognize the importance of the existence of so-called nonhomothetic preferences[88] for models of growth when he states that "… all recent models of economic growth share a defect which was characteristic of classical economy analysis. They have concentrated their emphasis exclusively on the technological side of economic reality and have entirely forgotten the other side, related to demand." (p. 69). He casts his critique in three propositions regarding the evolution of demand over time, all of which are important in the study of growth:

a) "… at any given level of per-capita income and at any given price structure, the proportion of income spent by each consumer on any specific commodity may be very different from one commodity to another;
b) as per capita real income increases, whatever the price structure may be, each *increment* of consumers' demand tends to concentrate on a particular group of goods and services. This group of goods and services gradually changes from one level of income to another. In other words, whatever the process through which consumers' preferences are reached they are such as to set up a definite order in the consumption of various goods, i.e. to saturate higher priority wants before passing to considering the satisfaction of further wants. Hence, as income increases, the tendency is not to increase proportionately the consumption of already bought goods and services, but rather to buy new goods and services or to satisfy old needs with different (and hopefully) better goods;
c) there is no commodity for which any individual's consumption can be increased indefinitely. An upper saturation level exists for all types of goods and services although at different levels of real income: it may be reached sharply - in the case of goods satisfying physiological needs - or only through a slow and long process as income increases - in the case of services yielding very sophisticated kinds of satisfaction - but its attainment is eventually inevitable. Demand for some particular goods (inferior goods)

[88] When preferences are nonhomothetic demand for products does not increase with the same amount for all products when income increases. That is, products do not all have the same (unitary) income elasticity and the composition of demand as measured by spending shares changes over time as income grows.

> may even decrease, after reaching saturation if better (superior) goods exist and if real income persistently goes on growing." (Pasinetti, p. 76/77, italics original).

Together these propositions imply the existence of a life cycle. With increasing income, demand for (new) products takes off in an attempt to satisfy lower priority wants, while ultimately total demand for a product is bounded by above due to saturation and over time demand for a single product, e.g. food, will lose its importance measured by its share in total demand (output) (cf. Engel's law)[89]. This life-cycle pattern is universal but the exact shape of it may vary over products and type of products. The final stage of the life cycle, however, is always that the share in total demand declines asymptotically to its final value. Of course for food, the final value is a positive (non-negligible) share, whereas black-and-white TV's will ultimately be completely replaced by color TV's. This in turn depends in part on the definition of a category, e.g. a broad category like food, or a narrow category like black-and-white TV's (as opposed to TV's in general) generally follows a different life-cycle pattern.[90] The model developed in the previous section is related to Pasinetti's work in several ways. The most important link is the emphasis Pasinetti places on the effect (rising) income has on consumer preferences and ultimately the composition of consumer demand. This type of consumer behavior is exactly what the previous section hypothesized underlies the absence of growth scale effects at the aggregate level while they may exist at the sectoral level, i.e. an intricate relation between consumer preferences, income, relative prices, technological progress, etc. An additional link is subsequently provided by the effect Pasinetti's propositions have on the evolution of demand over time, which is that products go through a life cycle with distinct phases of take off and ultimately decline.[91]

Combined, the remarks of Baumol and Pasinetti provide much needed credibility to the model of the previous section and its powerful assumptions. The intricate relation between consumer preferences, income, relative prices, technological progress etc., so often mentioned without much clarification of what it actually constitutes can now be further specified in a simple way. It can best be seen as some sort of feedback-loop between (productivity) growth and income:

---

[89] Engel already foresaw that "in the course of economic development agriculture would decline relative to other sectors of the economy." (Houthakker 1987a, p. 142).

[90] In Chapter 6, the life cycle and various approaches to this rather broad concept will be discussed at some length.

[91] In addition, an important difference between the model of the previous section and the models of growth (with or without scale effects) discussed in Chapter 2, besides the ones discussed in Paragraph 4.1.4, is only now fully apparent, namely the assumption regarding the homotheticity of consumer preferences. The models of Chapter 2 have generally used the standard Dixit-Stiglitz approach in which preferences are homothetic while it is now clear that the model developed in this chapter has implicitly assumed that preferences are nonhomothetic.

growth leads to a higher per-capita income which allows consumers to satisfy their wants with other (better) products and/or to satisfy lower priority needs through the consumption of new products (cf. Pasinetti). Of course new and/or better products have to be created first, for instance through product R&D, but this too is related to income because firms will only engage in this type of R&D when they find it profitable to do so, i.e. when a market for these new/improved products exists (cf. Sokoloff and Ades and Glaeser; see also footnote 86). For so far as rising income causes consumer demand to shift to goods produced in a (newly established) sector that is relatively unproductive (technologically stagnant), past growth has the ultimate effect of limiting future growth opportunities. Hence, technological progress broadly defined as increases in productivity and the introduction of new/improved products, in a way 'plants the seeds for growth's demise' and growth scale effects vanish very 'naturally'. In addition, the existence of a feedback-loop between technological progress in a broad sense and income gives new meaning to the term 'endogenous growth'.

Some studies have actually incorporated some sort of feedback-loop similar to the one discussed here in their models of growth, e.g. Laitner and Aoki and Yoshikawa. These models lend further support to the toy model developed in the previous section and consequently will be discussed next.

Laitner (2000)

Laitner presents a model in which a country's average propensity to save depends on its level of income. The basis for this is that saving is a product like any other and that demand for it follows an Engel-type curve. Laitner argues that this causes farmland to constitute a relatively large share of total wealth. He further argues that due to population increases and technological progress, the value of land increases over time. Households subsequently 'save' in the sense that their net worth increases as the value of their farmland rises. Engel's law, however, also predicts that with rising income demand shifts from agricultural goods to manufactured goods. In the production of manufactures, (farm)land is less important than (reproducible) capital and thus, with demand for food declining over time due to rising income, the capital gains on land will decline over time as well. If the overall willingness to accumulate wealth remains the same, this declining contribution of capital gains on land to wealth accumulation must be offset by an increase in the saving rate. Furthermore, the willingness and especially the ability to save may increase with rising income again leading to

a rise in the saving rate. All in all, causality does not only run from saving to income, as in the standard neoclassical model of growth (e.g. Solow 1956), but also from income to saving. That is, in Laitner's model, just like in the model developed in this chapter, demand for products, in this case 'saving', is intrically related to growth and income.

Aoki and Yoshikawa (2002)

In Aoki and Yoshikawa's model demand for a given product follows the typical S-shaped life cycle with a logistic demand curve as depicted in Figure 12.

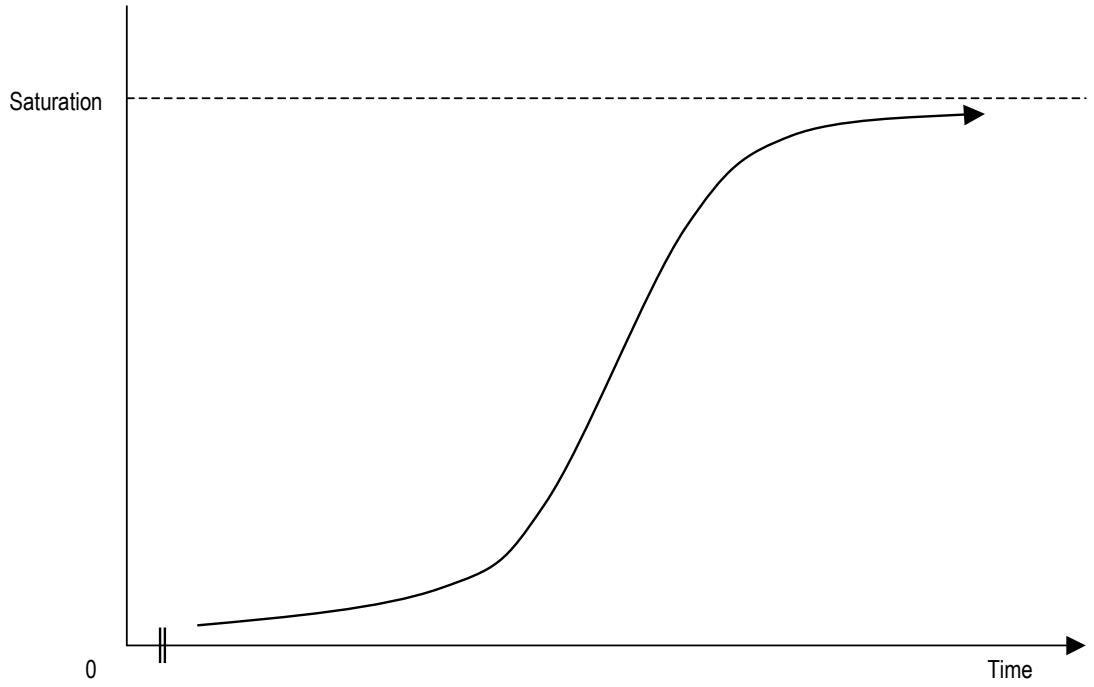


Figure 12, The typical S-shaped life cycle with a logistic demand curve (cf. Aoki and Yoshikawa).

As Figure 12 shows, demand for certain goods grows slowly in the beginning, then takes-off, and finally stagnates as it approaches the saturation point. In the Aoki-Yoshikawa model demand saturation is the fundamental factor restraining economic growth. This is at odds with the standard models of (R&D-based) growth in which growth is limited by diminishing returns to the accumulation of capital and/or by diminishing returns to the creation of new knowledge (R&D). From Engel's law it is apparent that demand for a certain product approaches its ceiling (or even vanishes completely) irrespective of the introduction of new products. Logistic growth of demand (i.e. the S-shaped life cycle) for a single product (or sector) has the logical consequence that growth at the aggregate level is high if the economy continuously introduces new products, which (temporarily) elicit high growth of demand. The

standard R&D-based models of growth (with or without scale effects) do allow for the creation of new products (and the destruction of old ones), but in these models the product life cycle is determined by the life cycle of the corresponding production technology. In the model developed by Aoki and Yoshikawa technological progress comes in the form of new products (or sectors) and has its impact on the growth rate of the aggregate through the creation of demand. While technological progress in the standard R&D-based models of growth comes in the form of an increase in total factor productivity or an outward shift of the production frontier, the growth effect of demand-creating innovation is not well captured by either of these interpretations of growth. Aoki and Yoshikawa use a simple AK-model (see for instance Romer 1986 and Lucas 1988), which has no diminishing returns to the accumulation of capital. The economy subsequently grows because resources are deployed - and capital is accumulated - for the production of goods, which enjoy rapidly growing demand. Growth is ultimately restrained by assuming that product innovation, the creation of new products, becomes harder over time (cf. Jones 1995b, Kortum 1997, and Segerstrom 1998), thus limiting the possibility of capital accumulation for the production of goods or sectors with fast growing demand. If in this model demand saturation would not take place, capital would continue to accumulate without the need of introducing new products, and growth would not be restrained.
In the model developed by Aoki and Yoshikawa consumer behavior (i.e. saturation and nonhomothetic preferences) play(s) an important role in ultimately restraining growth. They, however, still rely in part on a fishing-out effect to limit the introduction of new products, which is also necessary to restrain growth. One of the ideas of the model developed in the previous section conversely is that the introduction of new products is important because they steal away output from older products produced in sectors with a relatively high level of productivity. In this way, it differs drastically from Aoki and Yoshikawa's model, but still both models have many features in common, which lends credibility to the model of the previous section.

The goal of the above discussion of the work by Baumol (1967), Pasinetti (1981), Laitner (2000) and Aoki and Yoshikawa (2002), was to provide some credibility to the reality described by the simple model developed in the previous section. Indeed, where the assumption of an intricate relation between consumer preferences, income, relative prices, technological progress and so forth, causing demand to shift to (newly introduced) slow growing sectors (and ultimately causing the absence of growth scale effects at the aggregate

level), might have seemed far-fetched, it is actually based on ideas not all that uncommon in growth literature. In other words, it is not unlikely that saturation and nonhomothetic preferences (cf. Engel's law) combine with sectoral differences (cf. Baumol's disease) to cause the absence of growth scale effects (where logically such effects would be expected due to knowledge's nonrivalrousness).

This chapter continues with an examination of empirical evidence on structural change and sectoral growth differences and (some of) its determinants in the next section. This discussion will show that, as hypothesized in the model of Section 4.1, the sectoral structure of an economy indeed changes over time and that the various sectors have experienced widely differing rates of (productivity) growth. The underlying causes of these dynamics are examined on the basis of a study by Rowthorn and Ramaswamy (1999), which investigates the deindustrialization phenomenon, the shift in employment away from manufacturing towards services. Finally, these causes underlying structural change (e.g. deindustrialization) will be related to the issue of growth scale effects.

4.3 Structural change and deindustrialization

4.3.1 Data on structural change

It is a well-documented fact that in advanced countries, (the bulk of) gross value added and employment has shifted between sectors over time.[92] Figure 13 below depicts the output of ten sectors (comprising the whole economy), agriculture (Ag), mining (Mi), manufacturing (Ma), public utilities (PU), construction (Co), wholesale and retail trade (WaRT), transport and communication (TaC), finance, insurance and real estate (FIaRE), community, social and personal service (CSaPS) and government services (GS), as a percentage of total output (which equals demand in a closed economy) for the U.S. for the period 1950-1990.[93]

[92] See for instance Baumol, Blackman and Wolff (1989), Maddison (1991), van Ark (1996), Echevarria (1997), and Kongsamut, Rebelo and Xie (2001).
[93] See Table E.1 in Appendix E. Further note that in many studies on structural change attention is limited to three broad sectors, namely agriculture, manufacturing, and services.

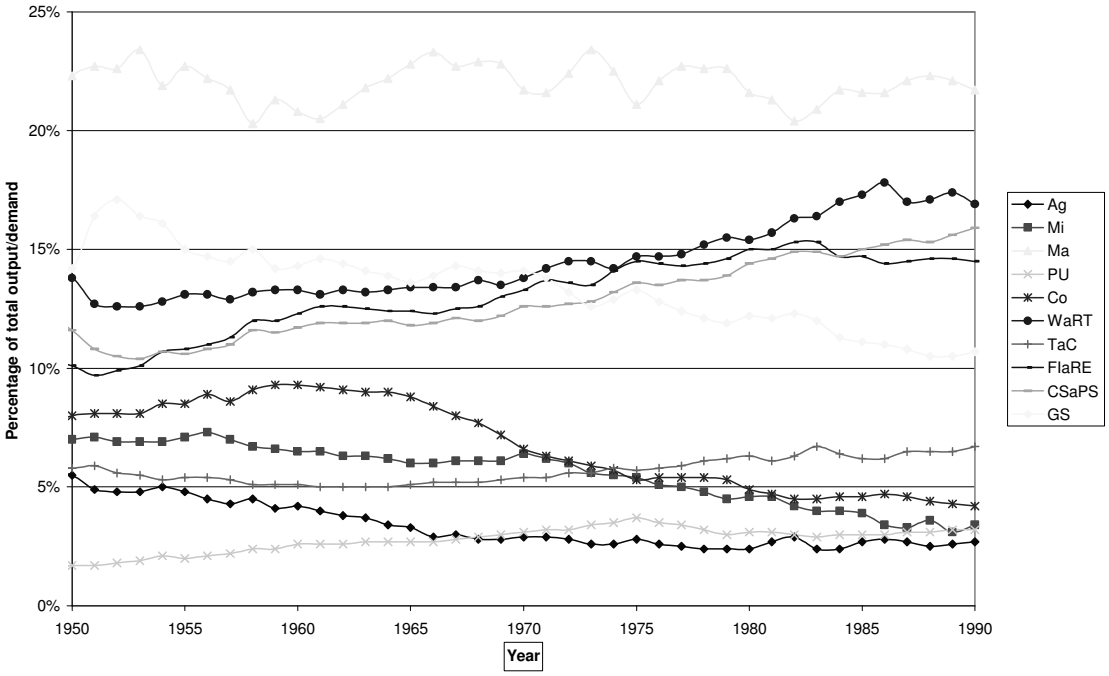


Figure 13, Sectoral structure of total output for the U.S. 1950-1990.[94]

Figure 13 shows that the sectoral structure of the U.S. is quite dynamic. During the period 1950-1990, some sectors have experienced a relative increase in demand for their products, some sectors have experienced a relative decrease in demand for their products, and some sectors have experienced a relatively steady demand for their products.

The logical next question is, of course, whether the growth rates of output in these sectors have differed much during this period. Table 4 below shows data on sectoral shares in total output and growth in output per worker for the sectors depicted in Figure 13.

[94] Source: van Ark (1996) and author's own calculations. See Table E.2 in Appendix E.

Table 4: Output shares and sectoral growth rates.[95]

Sector	Share in 1950	Share in 1990	Trend	Growth rate output per worker	Growth rate output
Ag	0.055	0.027	-6.85E-4	**2.85**	1.08
Mi	0.070	0.034	-1.01E-3	0.89	1.05
Ma	0.222	0.217	-9.37E-5	**2.43**	2.97
PU	0.017	0.032	3.50E-4	**2.99**	4.39
Co	0.080	0.042	-1.40E-3	-0.92	0.83
WaRT	0.138	0.169	1.23E-3	**1.52**	3.84
TaC	0.058	0.067	3.43E-4	**2.77**	3.60
FIaRE	0.101	0.145	1.29E-3	0.74	4.04
CSaPS	0.116	0.159	1.33E-3	0.63	4.04
GS	0.142	0.107	-1.33E-3	-0.10	2.00
GDP	...	...	...	1.20	3.01

The data shown in Table 4 has several interesting features. It shows that the output share of some high productivity growth sectors, that is sectors with an average annual growth in output per worker higher than the growth rate of GDP per worker (**bold**), has increased while the output share of other high productivity growth sectors has declined. Vice versa, the output share of some sectors with low rates of productivity growth has increased, while the output share of other sectors with low rates of productivity growth has declined. Despite the sectoral differences in growth in output per worker, structural change did not have a significant impact on the growth rate of productivity at the aggregate level. The growth rate of GDP per worker remained steady at about 1.20% a year throughout the entire period (1950-1990).[96]
The growth rates of sectoral output sketch a similar picture of widely varying sectoral growth rates but rather steady growth of aggregate output (GDP).[97] Of course, sectors that have experienced less than average growth of total (sectoral) output have experienced declining sectoral output shares (underlined).[98] Table 4 clearly shows this to be the case for agriculture, mining, manufacturing, construction, and government services.

4.3.2 Growth, trade, and deindustrialization

From the above-presented data, it is obvious that the sectoral structure of an economy is quite dynamic and that the sectors generally have different rates of growth. Even more interesting is

[95] Source: van Ark (1996) and author's own calculations. See Table E.2, Table E.3, Table E.5 and Table E.6 in Appendix E.
[96] The linear regression estimate of the annual growth rate of output per worker is 1.20% with a p-value of approximately zero (R^2 adj. = 0.9026). See also Table E.5 in Appendix E.
[97] The linear regression estimate of the annual growth rate of total output is 3.01% with a p-value of approximately zero (R^2 adj. = 0.9893). See also Table E.6 in Appendix E.
[98] Similarly, sectors which have experienced higher than average growth of labor productivity, i.e. agriculture, manufacturing, public utilities, wholesale and retail trade, and transport and communication, have of course experienced a decline in employment shares.

examining the causes underlying these observed dynamics. This is done on the basis of Rowthorn and Ramaswamy's study of deindustrialization, employment shifting from the manufacturing sector to the services sector.

In the literature on deindustrialization, several explanations for the observed relative decline of manufacturing employment have been offered, for example,

- that Engel's law plays a role in the demand for - and consequently output (production) of - manufactured goods: with rising income the share of expenditures on manufactured goods (eventually) declines (see Section 4.2; cf. Pasinetti 1981, Houthakker 1987a, and Clark 1951),
- differences in productivity growth at the sectoral level: productivity in the manufacturing sector grows fast relative to that in other sectors, so that over time less and less labor is required to produce a given amount of manufacturing products (see Section 4.2; cf. Baumol 1967),
- growth of north-south trade: imports from the developing countries are relatively low-skilled labor-intensive, e.g. manufactured goods, while exports from the developed countries are relatively skill-intensive, e.g. services.

Rowthorn and Ramaswamy try to explain deindustrialization for 18 advanced (OECD) countries over the period 1963-1994 as the result of several factors. They relate both the employment share, employment in manufacturing as a percent of total employment (EMPSHARE), and the output share, manufacturing value added as a percent of GDP (OUTSHARE), to a broad range of factors.[99]

The factors Rowthorn and Ramaswamy include are, the per-capita income (Y), the balance of manufacturing trade (TRADEBAL), i.e. total exports minus total imports, the manufactured import from less-developed countries (LDCIMP), the level of gross domestic fixed capital formation (FIXCAP), and the relative price (RELPRICE), i.e. the implicit price of manufactures divided by the implicit price of GDP.

The results of their regression analysis are depicted in Table 5 below.

[99] In the study of Rowthorn and Ramaswamy manufacturing is broadly defined and includes, amongst others, ISIC group 3 (see also footnote 93).

Table 5: Explaining deindustrialization.[100]

	ln(EMPSHARE)	ln(OUTSHARE)	ln(RELPROD)	ln(RELPRICE)
ln(Y)	11.78 (22.14)	6.32 (13.63)	0.30 (12.60)	...
$[\ln(Y)]^2$	-0.649 (22.39)	-0.350 (13.82)	...	...
ln(RELPROD)	...	...	...	-0.89 (28.97)
ln(RELPRICE)	...	-0.589 (18.37)	...	...
TRADEBAL	0.012 (7.11)	0.019 (12.96)	...	...
LDCIMP	-0.041 (3.79)	0.000 (0.01)•	0.085 (7.12)	-0.008 (0.85)•
FIXCAP	0.013 (8.68)	0.016 (11.71)	...	...
Country dummies exist?	Yes	Yes	Yes	Yes
R^2	0.90	0.84	0.88	0.80

Remarks: Absolute t-statistics in parentheses. (•) denotes insignificance at the one percent level.

As shown in Table 5, Rowthorn and Ramaswamy can explain a large part of the deindustrialization phenomenon. They find a statistically significant relation between the employment share and per-capita GDP, the balance of trade, import from less-developed countries (LDC's), and fixed capital formation. An increase in per-capita GDP, the balance of trade, and fixed capital formation leads to an increase in the employment share, while an increase of LDC import leads to a decrease of the employment share. Besides these variables, Rowthorn and Ramaswamy also find the square of GDP per capita to have a significant negative impact on the employment share. The relation between employment in manufacturing as a percent of total employment and income is thus hump-shaped. The maximum employment share is reached at a GDP per capita of about $9,000 (in 1986 purchasing power parity). Most OECD countries had reached this level of income by 1970, and some, e.g. the U.S., well before.

For the regression of the output share, which in light of the subject of this thesis is more important, similar results have been found. An increase in per-capita GDP, the balance of trade and fixed capital formation, lead to an increase in the output share, while an increase of the square of GDP per capita and the relative price leads to a decrease of the output share. The relation between manufacturing value added as a percent of GDP and income is thus also hump-shaped.

[100] Source: Rowthorn and Ramaswamy (1999).

From Rowthorn and Ramaswamy's results it seems that different levels of sectoral total factor productivity do not play a role in the observed dynamics of sectoral structure. As shown in the last two columns of Table 5, however, their work has found a highly significant negative relation relative productivity (RELPROD) and relative price: the higher relative productivity is, the lower the relative price of the (type of) product concerned is. Relative productivity in turn is strongly correlated to the per-capita income and the import from less-developed countries. The influence of import on relative productivity probably works through competition, domestically produced goods having to compete with imported goods, and the incentive this provides for innovation. On the other hand, per-capita income has a more complex relation to productivity. Over time, both the per-capita income and the relative productivity increase. Furthermore, a higher level of productivity generally leads to a higher level of income. Nonetheless, if some measure of time, e.g. the number of years elapsed since 1963, is included in the regression analysis, the per-capita income is no longer significant in explaining cross-country differences in relative productivity (see Rowthorn and Ramaswamy, p. 25).

The main conclusion of Rowthorn and Ramaswamy is that deindustrialization can largely be explained by the interactions between changing demand patterns, i.e. demand moving away from manufactures towards services, and differences in productivity growth, the manufacturing sector grows faster than the services sector, which lead to a decline in the relative price of goods produced in the sector with the highest rate of productivity growth (manufacturing). These causes of deindustrialization, internal to an economy, can explain about 80% of the observed relative decline of manufacturing employment, while external factors like north-south trade can explain approximately 20% of the relative decline.

The results of Rowthorn and Ramaswamy strengthen the idea that there is an intricate relation (a feedback-loop) between technological progress (measured as productivity growth) and income and that this relation is to no small extent brought on by consumer preferences (demand). Hence, the conclusion from this discussion of explanations of deindustrialization echoes that of the previous sections; consumer preferences, income, relative prices, R&D (technological progress) etc., are highly intertwined and work together to cause growth scale effects to vanish (at the aggregate level).

In addition, Rowthorn and Ramaswamy's finding of a hump-shaped relation between the output share (and the employment share) and per-capita income might provide a link between structural change (deindustrialization), and scale effects in growth as will be discussed in the

next paragraph.

4.3.3 Structural change and growth

As mentioned in Chapter 2, Papageorgiou, investigating the issue of scale effects in growth, finds a hump-shaped relation in a cross-country regression of growth in per-capita GDP and the level of human capital. Since the level of human capital is undoubtedly closely related to the level of per-capita income, the hump-shaped relation Papageorgiou finds between growth in per-capita GDP and the level of human capital (as opposed to the level of per-capita income) is very similar to the hump-shaped relation Rowthorn and Ramaswamy find between the share of manufacturing in total output and the level of per-capita income (as opposed to the level of human capital) (see the previous paragraph). Hence, there may be a link between structural change (deindustrialization) and the existence of scale effects in growth (of total output). This link is further clarified in Table 6.[101]

Table 6: Structural change and growth.

Study	Dependent variable	Independent variable(s)		Relation
Papageorgiou	Growth	Human Capital	Human Capital^2	Hump-shaped
Rowthorn and Ramaswamy	Manufacturing output share	Income	Income^2	Hump-shaped
…	Human capital	Income		Linear
…	Growth	Income	Income^2	Hump-shaped
…	Growth	Manufacturing output share		…

To test the logic depicted in Table 6, here several of the relations depicted in this table are investigated (again) using data from Jones (1998b) on, amongst others, the level of human capital , the annual growth rate of GDP per worker, and the level of GDP per worker relative to that of the U.S., for 109 countries (see Table F.1 in Appendix F).[102] The relations that will be investigated are relations between:

[101] Note that the interest here is not on human capital, but rather on the similarity of the hump-shaped relation, which Papageorgiou relates to the issue of growth scale effects, while in the study of Rowthorn and Ramaswamy it plays a role in the sectoral structure of economies.

[102] Jones' (1998b) data set, available at http://emlab.berkeley.edu/users/chad/TableC2.asc (d.d. December 2003), contains only cross-sectional data so that the segmented time trend test of Papageorgiou (2002) cannot be repeated. In addition, his data set does not contain a measure of per-capita GDP. GDP per worker, however, is a measure closely related to per-capita GDP.

- the growth rate of GDP per worker (proxying for the growth rate of per-capita income) and a measure of human capital (the average educational attainment in years) (cf. Papageorgiou),
- the growth rate of GDP per worker and the level of human capital within four income classes (cf. Papageorgiou),
- the level of human capital and relative GDP per worker (see Table 6),
- the growth rate of GDP per worker and relative GDP per worker (see Table 6).

The results of this investigation are summarized in Figure 14 below (see also Table F.3, Table F.4 and Table F.5, in Appendix F).[103]

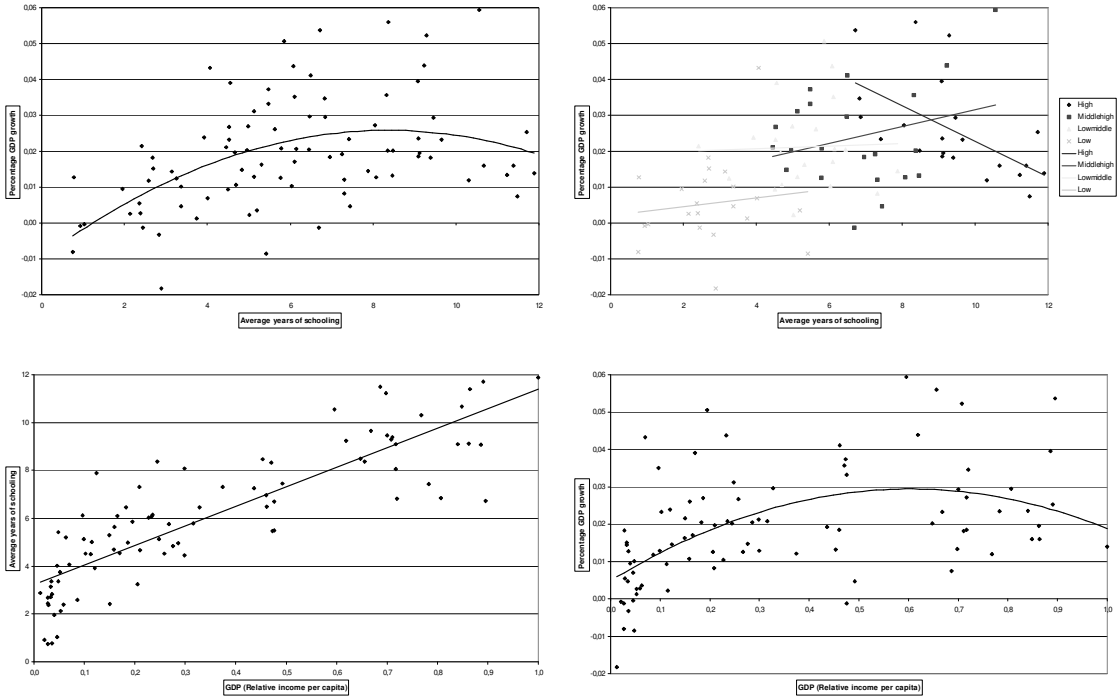


Figure 14, Growth, schooling, and income.[104]

More concrete, the results of the investigation of these relations are that:

- There is a (familiar; see Papageorgiou 2002) hump-shaped relation between the growth rate of income, as proxied by the growth rate of GDP per worker, and Jones' (1998b)

[103] Again note that the interest here is not on human capital. The regressions with human capital are only needed to clarify the link between Papageorgiou's finding, when investigating the issue of growth scale effects, of a hump-shaped relation between growth in per-capita GDP and the level of human capital (as opposed to the level of per-capita income) and Rowthorn and Ramaswamy's finding, when investigating deindustrialization, of a hump-shaped relation between the share of manufacturing in total output and the level of per-capita income (as opposed to the level of human capital).

[104] Source: Jones (1998b). Due to missing data, the original data set was reduced to 88 (minus 21) countries. See also Table F.1 in Appendix F.

measure of human capital, the average educational attainment in years. The relation fits the data rather well, all coefficients are significant at the one percent level and some 22% of the total observed variance in growth of GDP per worker is explained (see Table F.3 in Appendix F);

- There is a statistically significant relation between the growth rate of GDP per worker and the level of human capital within four productivity (income) classes only for the class with the highest level of GDP per worker (relative to the U.S.): the level of human capital negatively affects the growth rate (see Table F.3 in Appendix F). Papageorgiou conversely finds a statistically significant relation only for the lowest two income (productivity) classes: the level of human capital negatively affects the growth rate of per-capita GDP. An explanation for these differing results may be the way the income classes are constructed (see Table F.2 in Appendix F);
- Most of the variance in the level of human capital in a cross-section of countries can be explained by corresponding variances in GDP per worker (see Table F.4 in appendix F);[105]
- A hump-shaped relation is highly significant in explaining cross-country variances in growth of GDP per worker from variances in GDP per worker (see Table F.5 in Appendix F).[106]

The conclusion from these findings is that the logic depicted in Table 6 is confirmed. There probably is a relation between structural change (deindustrialization) and growth (with or without scale effects), as both are related to income in a similar matter.

Hence, to conclude this paragraph and section, data on sectoral employment shares (i.e. structural composition) is used to explain growth experiences over several episodes.[107] Maddison (1991) provides data on the employment share of the agricultural sector (the share of agricultural employment in total employment), the manufacturing sector and the services sector respectively, for 1870, 1950, 1973 and 1987, for 13 advanced (OECD) countries. In

[105] In addition, a quadratic regression analysis was applied. This analysis revealed a familiar hump-shaped relation between the average educational attainment in years and GDP per worker reveals. This hump-shaped relation can be seen as depicting the income elasticity of demand for education.

[106] Table F.5 in Appendix F also shows the result of applying a quadratic regression analysis to explain the investment share of as a function of income. A significant relation is found, although a linear relation can explain the observed variances almost equally well. The finding is nonetheless supportive of Laitner's study in which a bi-directional relation between saving and growth (and ultimately income) exists.

[107] See also studies by Baumol, Blackman and Wolff (1989) and Echevarria (1997), which provide findings similar to that of Papageorgiou (2002) and relate them to sectoral structure and in the case of Echevarria construct a model to explain the findings using nonhomothetic preferences.

addition, Maddison provides data on output per man-hour for the same years and the same countries. From this data, the growth rate of output per man-hour can be calculated for three periods, 1870-1950, 1950-1973, and 1973-1987, for each of these countries, the results of which are shown in Table G.1 in Appendix G. The 39 observations (13*3) are subsequently used in a regression analysis in which differences in the sectoral structure are used to explain the differences in growth rates across periods, e.g. 1870-1950, 1950-1973 and 1973-1987, and across countries.

The analysis shows that variance in both the agricultural and the manufacturing employment share can partly explain observed variance in growth rates and it seems that the higher the share of these two sectors in total employment, the higher the growth rate (see Table G.2 and Table G.3 in Appendix G). This is in support of the idea of Baumol that technologically stagnant sectors like services become the dominant economic sector over time. The explanatory power of either the level of industrialization (manufacturing share) or the sectoral structure (agricultural and manufacturing share[108]) regarding growth differences is, however, limited. In fact, dummies denoting the period under investigation are most significant in explaining growth differences over the various periods. Nonetheless, a typical regression result is: p-value share agriculture = 0.123, p-value manufacturing share = 0.018; R^2 adj. = 0.108, F-statistic = 3.29, Significance F-statistic = 0.0486 (see Table G.2 and Table G.3 in Appendix G).[109] In addition, future work using extended data sets may be more able to use sectoral structure in explaining growth experiences and possible relations between size (suitably measured) and rates of growth (cf. Baumol, Blackman and Wolff 1989, Echevarria 1997, and Papageorgiou 2002).

4.4 Discussion and conclusions

To conclude this chapter, this section starts with summarizing important findings and results of this chapter regarding the issue of growth scale effects and more specifically important aspects of economic reality related to it.

The most important finding is not so much a finding as it is a premise, namely that there is

[108] In the various regressions that use the agricultural- and the manufacturing share as explanatory variables, the sectoral employment share of services is of course implicitly taken into account.
[109] Besides the work reported in Table H.2 and Table H.3 experiments with other explanatory variables, e.g. services share, have been conducted, which yielded similar results as those depicted in these Tables.

some sort of feedback-loop between factors like consumer preferences, income, relative prices, technological progress and so forth, i.e. consumer preferences, income, relative prices, technological progress and so forth are mutually related, and that this causes demand to shift away from (older) relatively productive sectors towards (younger) relatively unproductive sectors (and ultimately causing the absence of growth scale effects at the aggregate level). The model, or mathematical stylization, of Section 4.1 assumes a feedback-loop between these factors to reconcile the seemingly contradictory facts that growth at the aggregate level does not exhibit a scale effect while it may exhibit such an effect at the sectoral level. The mechanism underlying this result is simply the continuous shift of demand to new sectors in which the level of productivity is relatively low.
Where the assumption of an intricate relation, i.e. a feedback-loop, between consumer preferences, income, relative prices, technological progress etc., seemed far-fetched this depiction of reality is actually supported in work discussed in Section 4.2. Mainly based on Engel's finding of food expenditures declining as a share of total income as total income rises, various studies (including of course some of the ones discussed in Section 4.2) assume that consumption of a good is subjected to saturation and that preferences are nonhomothetic (as opposed to homothetic; see for instance Dixit and Stiglitz; see also Pasinetti). In addition, the toy model presented in this chapter is very similar to the model of unbalanced growth presented by Baumol, in that in both models consumer behavior and sectoral growth differences (i.e. technological opportunity) play an important role. In Baumol's model these two factors are important for its prediction that it becomes more difficult over time for the economy to maintain its overall rate of output growth (p. 421), while for the simple model presented in this chapter, these factors are important causes underlying the absence of growth scale effects.

Empirical support for the existence of an intricate relation between consumer preferences, income, relative prices, technological progress etc., is subsequently provided not only by Engel's law, which is one of the most robust findings in economics (cf. Houthakker 1987b and Deaton and Muellbauer 1983), but also by empirical work that has investigated the deindustrialization phenomenon (see Section 4.3). This work finds that the observed decline in the share of manufacturing in total employment and output is mainly the result of factors internal to an economy (as opposed to external factors like north-south trade). 80% of the observed deindustrialization can be explained by the interactions between changing demand patterns, i.e. demand moving away from manufactures towards services, and sectoral

differences in productivity growth, productivity in the manufacturing sector grows faster than in the services sector causing a decline in the relative price of products produced in the fastest growing sector (manufacturing).

Following the results of this empirical work on structural change, sectoral structure (and changes therein) was used to explain observed differences in growth. This did not lead to much insight regarding the issue of growth scale effects but sectoral structure was found to matter in explaining observed growth differences. In addition, future work that is more thorough may shed more light on the relation between sectoral structure and growth with or without scale effects.

Overall, the empirical work on structural change (deindustrialization) confirms the idea forwarded in Sections 4.1 and 4.2 especially, that there is some sort of complex interaction between (productivity) growth and income.

Final evidence supporting the model developed in this chapter actually comes from the findings of previous chapters and Chapter 3 in particular. The model originally set out to provide a mathematical description of an intuitive mechanism of how scale effects vanish at the aggregate level while they may exist at the sectoral level. In doing so, however, it actually incorporates some of the ideas forwarded in Chapter 3, which in turn lends further credibility to the correctness of its depiction of economic reality:

- life-cycle effects are important. Demand changing from older to newer sectors causes products to go through a life cycle in which demand for the product takes off and grows almost exponentially (e.g. the take-off phase), reaches its maximum, and ultimately declines asymptotically to zero;
- dilution of knowledge through technological distance, which limits spillovers. The assumption of productivity in new sectors being lower than in older sectors implies that the knowledge is partially product- or sector-specific. This product-specific knowledge is subsequently diluted when applied to the production of other products than that for which it was originally conceived. That is, knowledge only reaches its full productive potential when used in the production of the products for which it was originally conceived, and reaches only part of this potential when used in the production of other products. What is more, in the model presented in Section 4.1 the concept of dilution is not used to limit intertemporal (dynamic) spillovers, but, as is consistent with the original evidence of Adams and Jaffe, to limit static spillovers, which have a level effect.

In this way, the model developed in Section 4.1 is related to the idea of endogenous network participation developed in the previous chapter for it describes a mechanism by which network participation may be endogenous instead of random while spillovers are still limited to a sufficient extent to cause the absence of growth scale effects.

Gathering clues for the understanding of - and the ability to credibly model - growth without scale effects, the conclusion of this chapter subsequently is simply that more attention should be paid to the intricate relation that exists between factors important in growth, e.g. consumer preferences, income and technological progress. If anything, this chapter has shown that the intricacies associated with these factors and their mutual relations, are important in the study of growth in general and that part of it concerned with scale effects in particular. For instance, until now, the number of models incorporating the robust fact of structural change and sectoral differences is limited (with Echevarria 1997, Kongsamut, Rebelo and Xie 2001 and Aoki and Yoshikawa 2002 as notable exceptions).

Nonetheless, an important critical remark must also be made. The incentive consumers have to shift demand from one good to another good is of course that they find that the new good is better able to satisfy their wants than the old good; shifting demand will increase their well-being. By completely focussing on output, the model has neglected the welfare or well-being aspect of the model and economic reality, which is related to income (output) but in which other factors, e.g. product variety and consumer surplus (typically associated with the area under the demand curve), also play an important role. Indeed, growth of well-being (or welfare) in the model may exhibit a scale effect; the combined effect of the continuous introduction of new goods which provides consumers with more opportunities to choose from and a possible increase in consumer surplus may lead the rate of growth of well-being to exhibit a scale effect.
The well-being effects of the mechanism driving the result of the model are not accounted for by the model however. This is due the fact that the model is very simple and has not explicitly modelled the demand side with (relative) prices etc. Because the issue of well-being is important nonetheless, also for the idea of saturation and nonhomothetic preferences, it, and especially its relation to income, is discussed at some length in the next chapter.

To conclude the chapter and this section, some interesting directions for future work are presented.

- A direction for theoretical work might be to scale-down a model along the lines of the one presented in this chapter and focus on demand shifts pertaining to a single product rather than to entire sectors. Just like sectoral growth may exhibit a scale effect while growth at the aggregate level does not, the growth rate of a single product within a certain sector may exhibit a scale effect while growth at the sectoral level does not.
- An interesting direction for empirical work is of course to investigate whether scale effects are existent in the other sectors, agriculture and services, and/or in the production of a single product.

5 Measurement issues in the study of R&D-based growth

At several points in this thesis, the issue of measurement has come up. The simple model developed in the previous chapter has made plausible an intuitive reason why growth (at the aggregate level) does not exhibit a scale effect, but has not addressed the issue of well-being in which not only income is important but also for example the variety of products available for consumption. In Chapter 2, it was assumed that knowledge can be cardinally measured. This was done to address up front possible critique like that of Steedman (2001) (far-fetched as it may be). Finally, also in Chapter 2, the trouble of accurately accounting for quality changes and the introduction of new products in measures like the Consumer Price Index (CPI) and (consequently) GDP was mentioned. This combined with the fact that looking at measurement issues is useful in itself provides the raison-d'être of this chapter. More specifically this chapter aims at determining the role measurement issues (in a broad sense) could play in the issue of growth scale effects and maybe even gathering clues pointing to important aspects of economic reality that could play an important role in explaining the absence growth scale effects where logically such effects are expected as the result of knowledge's nonrivalrousness.

For this, in the first section (5.1) the concept of knowledge is elaborated on by specifying further what in this thesis (and models of growth) is actually meant by 'knowledge' in Paragraph 5.1.1 and by discussing problems related to the (economic) measurement of knowledge in Paragraph 5.1.2. By doing so, the section points out certain things that seem important with regards to models of growth (with or without scale effects) and the way knowledge is handled in them. Section 5.2 subsequently discusses the measurement of income which is related to the measurement of the (true) cost of living in which, in turn, quality change and the introduction of new products, play an important role. As both income and the cost of living are related to well-being, well-being plays a large role in the issue of growth without scale effects with some remarkable implications. These two main issues are discussed in Paragraph 5.2.1 and Paragraph 5.2.2 respectively. In Section 5.3, a discussion and some conclusions are presented. Special attention is paid to pointing out the relevance of certain findings to the pervading issue of scale effects in growth.

It is important to note beforehand, that at times, this chapter will be rather 'philosophical' in nature and each (partial) answer seems to evoke yet another question. What is more, this chapter will not provide a comprehensive discussion of the issue of measurement in the study of growth simply because such a treatment of the issue is well beyond the scope of the present thesis, if at all possible. This is reflected in the sparse use of references, where at times, especially regarding the relation between income and well-being, literally hundreds of studies have presented interesting findings that are of interest to the issue.[110] All this is not to say, however, that the usefulness and importance of several of the issues raised in this chapter for either the study of growth in general or the issue of growth scale effects in general, is limited or strongly diminished by its narrow scope.

5.1 TFP and the knowledge stock

Steedman fulminates against many contributions to what he labels "New (Endogenous) Growth Theory", that is models of growth in which, following Romer (1986), growth is 'endogenous'. His main point of critique is that these models, e.g. Romer (1986 and 1990) and Jones (1995b), make reference to knowledge, or 'a stock of knowledge', without making clear what is meant by it. Steedman finds, for instance, that the relevant mathematical equations in these contributions are meaningless unless knowledge can be cardinally measured and/or otherwise be seen as a productive resource like any other, e.g. capital or labor. In addition, he states that the new growth theory is "… theory without the minimal conceptual clarity required to make that theory worthy of attention." (p. 3). Although Steedman's critique is rather straightforward and can be applied to a large array of subjects in economics (e.g. "What is capital?"), it does point to the need for further elaboration on (i) the way knowledge is used in models of (R&D-based) growth, (ii) the measurement of knowledge.

[110] An important reference that nonetheless must be made before the actual start of this discussion of measurement issues in the study of R&D-based growth is to the appendix to Chapter 12 of Aghion and Howitt (1998). In their treatise on endogenous growth, which is vastly developing into a standard work on this subject, several of the issues regarding measurement in the study of growth raised in this chapter, are dealt with.

5.1.1 Definition and use of knowledge in R&D-based models of growth

In the study of economic growth, knowledge is regarded as an input to the production process. Take for instance the standard production function from the neoclassical model in the familiar Cobb-Douglas notation (e.g. Solow 1956):

$$Y = AK^{\alpha}H^{1-\alpha} \tag{5.1}$$

where Y is total output, A represents the knowledge stock, K is the amount of capital, L is the amount of human capital (or effective labor), and α is some number between 0 and 1. This equation offers a simple framework for studying growth. It shows that growth, i.e. an increase in total output, can have three sources, (1) an increase in knowledge, (2) an increase in capital, and (3) an increase in effective labor (human capital).

It is easy to think of ways in which the amount of capital and human capital used in the production process can be increased. An increase in saving and a corresponding increase in investment would, abstaining from depreciation, increase the amount of capital available for production. Similarly, an increase in the participation of people (previously inactive) in the production process would increase total output as would making the current labor force more effective (e.g. through education). In addition, it can easily be seen that the production function depicted in equation (5.1) exhibits constant returns to scale in the input of capital and effective labor. If the input of capital and human capital is doubled, total output will double as well. Likewise, the production function exhibits diminishing returns to a single production factor.

For the case of growth resulting from an increase in the stock of knowledge, the matter is less straightforward. The main point of the R&D-based models of growth (or endogenous growth models or Schumpeterian growth models (see footnote 1)) is that an increase in the stock of knowledge, and ultimately economic growth, is the result of profit maximizing behavior of economic subjects. However, here the critique of Steedman hits upon an important point, which is that in models of endogenous growth knowledge "… is far from being a crystal clear concept." (p. 2). To be able to think of more concrete ways in which the knowledge stock can be increased it is necessary to have a clue of what it is that constitutes knowledge. The definition of knowledge as information available to a person (or persons) that determines his (their) ability to combine resources to produce goods (see Chapter 2), does clarify the concept slightly. This definition, however, is not precise enough to view knowledge like a productive resource like any other. Consider for instance the statement above that the production function

depicted in equation (5.1) exhibits constant returns to scale in the input of capital and effective labor. For knowledge as defined above, such a statement still cannot be made, simply because the knowledge stock may very well contain items that do not lend themselves for replication as capital and human capital do. Hence, what is needed is further specification of the items that make up knowledge.

Here it is tried to find such a specification by combining the result of (early) growth accounting exercises with (contemporary) growth regressions. In growth accounting, a production function like that depicted in equation (5.1) is used to account for observed growth by attributing it to growth in the input of capital and effective labor.
To see how this works, equation (5.1) is rewritten to depict growth rates:

$$\hat{Y} = \frac{\dot{Y}}{Y} = \hat{A} + \alpha\hat{K} + (1-\alpha)\hat{H} \tag{5.2}$$

Equation (5.2) clearly shows that observed growth ($\hat{Y}$) can be accounted for by growth of the inputs of the production function. In general however, early growth accounting exercises tried to explain observed growth only from growth of capital and (effective) labor inputs and were subsequently left with a (large) unexplained residual (see for example Abramovitz 1993 and Griliches 1996). This residual was generally viewed as an increase in total factor productivity (TFP). In terms of the production function of the neoclassical model depicted in equation (5.1) an increase in TFP of course equals an increase in the knowledge stock (technological progress), as can easily be shown by rewriting equation (5.2):

$$\hat{A} = \hat{Y} - \alpha\hat{K} - (1-\alpha)\hat{H} \tag{5.3}$$

Abramovitz (1993), however, notes that "… the residual in its early primitive form was a cover for many sources of growth besides technological advance […] [S]o the Residual was really a grab-bag." (p. 218).
Because the sort of knowledge that is of interest in this thesis, is the one that is associated with technological advance, the knowledge stock is part of the residual together with other sources of growth, and (early) growth accounting provides only a starting point for the further specification of knowledge. This is where growth regressions are useful because they provide exactly this. Recent growth regressions have collected a broad range of variables that are significant in explaining growth. Sala-i-Martin (1997) for instance has found that the absolute latitude (distance to the equator) and a dummy variable indicating whether a country is a

former Spanish colony have a significant positive relation and a significant negative relation respectively to observed growth. The effect on growth of the variables found by Sala-i-Martin especially those relating to geography, the political situation and religion, naturally cannot be attributed to capital and/or human capital and therefore will show up in the residual, i.e. in TFP. Hence, the knowledge stock when defined as TFP is a function of a many variables including the aforementioned and of course past R&D effort:

$$A = f(\text{Geography, Politics, Religion, R\&D, etc.}) \tag{5.4}$$

The knowledge stock as defined by equation (5.4), contains variables, like the absolute latitude, that obviously do not lend themselves for replication like capital and effective labor (they are non replicable). In addition, this definition does not match the previous one of knowledge as information available to a person (or persons) that determines his (their) ability to combine resources to produce goods. The absolute latitude can hardly be characterized as information determining a person's productivity.

What is needed to specify knowledge more detailed so that it can be seen as a productive resource like any other consequently is to divide the knowledge stock as defined by equation (5.3) and (5.4), the residual, in two parts:

1. The first part is what tentatively can be named the external productivity part and contains variables that determine an economic subject's productivity, but do not lend themselves for replication, and/or are not likely affected through actions of profit maximizing economic subjects (non-replicable).[111] External productivity is determined by variables like the absolute latitude which do not change over time, and also variables like political variables and civil liberties which can change (increase or decrease) over time just not as the result of profit maximizing behavior.
2. The second part of the residual then is the part that contains variables that do lend themselves for an increase resulting from the actions of profit maximizing economic subjects (replicable). This part is knowledge in the sense that it is consistent with the definition of knowledge as information available to a person (or persons) that determines his (their) ability to combine resources to produce goods. In addition, this part of the residual is the knowledge stock or 'A' that is the central theme in the new (endogenous) growth theory (see for example the survey in Jones 1999). Steedman's critique is

[111] The idea that the knowledge stock results from - and changes as a result of - the actions of profit maximizing economic subjects is of course the defining property of the new endogenous growth theory as opposed to the neoclassical theory of growth (e.g. Solow 1956). Further note that the external productivity part of TFP comprises all knowledge that has not been created as the result of deliberate actions by economic subjects.

subsequently justified in that most contributions to new (endogenous) growth theory do not explicitly state that their use of the term the 'knowledge stock' refers to the total knowledge accumulated through the actions of profit maximizing economic subjects. On the other hand, from an equation like equation (2.2) which depicts the knowledge production function, it can be deduced that the knowledge stock in the new (endogenous) growth theory refers to knowledge as defined here, i.e. as information which results from the maximizing behavior of economic subjects and when available to an economic subject partly determines its ability to combine resources to produce goods.[112]

The distinction made above between external productivity, which still very much resembles a grab-bag (Abramovitz 1993), and R&D-based knowledge is undoubtedly a useful one. However, further specification is still required, since this distinction brings forth new questions, especially regarding the way in which external productivity and knowledge interact with each other and other resources (capital and effective labor) to determine total output. That is, how should the production function be adjusted, if at all, to incorporate both the knowledge stock and external productivity.

The most logical starting point for answering this question is that it should be possible to overcome a disadvantage in external productivity resulting from for instance being a former Spanish colony by additional R&D effort. On the other hand however, if at some point, knowledge is diffused perfectly, countries might still differ in their level of per-capita income. For example, a country that develops a new breed of crops that performs well under the country's climatic conditions, may in the long run not be able to compensate for their climatic disadvantage, because the same crop may be used in other countries to create even more output. In other words, in the long run, convergence may very well be 'conditional'.

Taking the above into account, here two straightforward reformulations of the production function depicted in equation (5.1) are proposed:

$$Y = (E * A) \cdot K^{\alpha} H^{1-\alpha} \tag{5.5}$$

$$Y = (E + A) \cdot K^{\alpha} H^{1-\alpha} \tag{5.6}$$

[112] To paraphrase, this second part of knowledge defined as TFP, can be referred to as R&D-based knowledge, as opposed to non R&D-based knowledge which is captured in the first part, i.e. in the external productivity part (see also footnote 111).

with A the (R&D-based) knowledge stock and E external productivity.[113] It is obvious that in either of these specifications of the production function a disadvantage in external productivity can be overcome. In the newly proposed production function specifications, growth can result from four sources, namely from the traditional increases in capital (K), effective labor (H), and/or the knowledge stock (A), and from an increase in external productivity (E). There is, however, also a very important difference between the two reformulations regarding the extent of the effect a similar-sized increase ($\kappa > 1$) in all the inputs, i.e. in E, A, K and L, has on output and total factor productivity (TFP(E,A)) and output (Y(E,A,K,L)). As discussed in Chapter 2, the replication argument implies that there are constant returns to scale in the rivalrous inputs, $Y(E,A,\kappa K,\kappa L) = TFP(E,A)\cdot F(\kappa K,\kappa L) = TFP(E,A)\cdot\kappa F(K,L) = \kappa Y(E,A,K,L)$ (see also footnote 113), but increasing returns to the rivalrous and the nonrivalrous inputs combined (see also Romer 1990), and on this point the two proposed reformulations of the production function do not differ. The situation changes drastically, however, when all the productive inputs experience the same $\kappa > 1$ increase. The result of these increases then is, either:

a) $Y_{5.5}(\kappa E,\kappa A,\kappa K,\kappa L) = TFP_{5.5}(\kappa E,\kappa A)\cdot F(\kappa K,\kappa L) = (\kappa\cdot\kappa)\cdot TFP_{5.5}(E,A)\cdot\kappa F(K,L) = (\kappa\cdot\kappa\cdot\kappa)\cdot Y_{5.5}(E,A,K,L)$ (based on equation 5.5), or

b) $Y_{5.6}(\kappa E,\kappa A,\kappa K,\kappa L) = TFP_{5.6}(\kappa E,\kappa A)\cdot F(\kappa K,\kappa L) = \kappa TFP_{5.6}(E,A)\cdot\kappa F(K,L) = (\kappa\cdot\kappa)\cdot Y_{5.6}(E,A,K,L)$ (based on equation 5.6).

Hence, the effect an R&D-induced increase of the knowledge stock (A) has on the level of TFP and output depends on the exact formulation of the production function (Y(E,A,K,L)) or more specifically on the exact formulation of TFP as a function of E and A (i.e. TFP(E,A)). This in turn has consequences for the definition of knowledge (A) and the previous definition of knowledge can be made slightly more exact, although two different definitions, each corresponding to one of the two extreme formulations of the production function depicted in equation (5.5) and (5.6) respectively, are needed.

1. Knowledge is information which results from the maximizing behavior of economic subjects and when available to an economic subject determines its ability to combine resources to produce goods within the confines of an external productivity parameter (this

[113] A much more general form of the production function in which distinction is made between the two different parts, external productivity and R&D-based knowledge, making up TFP would be: $Y(E,A,K,L) = TFP(E,A)\cdot F(K,L)$. The two reformulations depicted in equation (5.5) and (5.6) respectively should be seen as extreme forms of this general specification, serving only the purpose of providing insight.

is the definition that follows from the first specification of the production function in equation 5.5).

2. Knowledge is information which results from the maximizing behavior of economic subjects and when available to an economic subject - together with an external productivity parameter - determines its ability to combine resources to produce goods (this is the definition that follows from the second specification depicted in equation 5.6).

Disappointingly, there is thus not a single definition of knowledge that can be used throughout this thesis. On the other hand, the subtle difference between the two definitions of knowledge and especially the difference between the two extreme specifications of the production function play a role in (i) the issue of growth scale effects, and (ii) the reconciliation of the findings of Porter and Stern (2000) with those of Jones (1995a) (see Chapter 3), as will be discussed in the concluding section (Section 5.3).

5.1.2 On the measurement of knowledge

The adjusted definition of knowledge given above has made the concept of knowledge clearer, but unfortunately (yet unsurprisingly) new questions arise, for instance regarding the way knowledge and the knowledge stock is measured. Of course this a question that cannot be discarded easily, because knowledge is at the very core of R&D-based models of growth (with or without scale effects) and the factor most central in this thesis. Keeping in mind that it is probably impossible ever to make such a complicated concept as knowledge crystal clear, this paragraph will thus try to formulate some (partial) answers to this question (see also Machlup 1980).

Generally, there are two, distinct though closely related, ways to measure knowledge (i.e. its value), (1) by measuring its effect on productivity, and (2) by measuring the costs of creating the knowledge. These two methods of measurement and the problems surrounding their practical usefulness will be discussed here.

1. Starting-point for the first method is that if a person somehow has ‘more knowledge’ at his ‘disposal’ (relative to some reference amount of knowledge), his ability to produce goods has increased and subsequently the amount of extra knowledge available can be measured by the extra output (per unit of productive resource) it helps produce. As logical as this method of measuring knowledge, i.e. by determining the effect it has on production, seems, it, like the other method discussed below, is surrounded with problems. One such problem is immediately clear. Assume for instance that someone

thought of a new way to make wooden shoes, allowing him to double his daily production from one to two wooden shoes, whilst still using only one chisel and one unit of labor (i.e. the amount of capital and labor used in the production process have remained constant). Now it would still be impossible to 'measure' the amount of this knowledge. The amount of knowledge depends on the absolute amount of production to which it is applied. If after a year the woodworker decides to try his luck as a farmer, the knowledge will have helped to produce 365 extra wooden shoes. The extra amount of knowledge subsequently equals 365 wooden shoes, or rephrasing to more economic terms, the value of the extra knowledge is 365 wooden shoes. Note, however, that the use of value immediately raises a second important problem associated with using productive use as a measure of knowledge, namely that with the increased output of wooden shoes, the value of a single wooden shoe, as compared to other products, might have declined. For example, where previously 5 pairs of wooden shoes bought one chisel, now 10 pairs of wooden shoes are required to obtain one chisel.[114] Measured in chisels then, the extra knowledge has not increased output, because total output produced using this extra knowledge still equals 73 (chisels). The value of the extra knowledge is subsequently nil when measured in chisels. Similarly, the extra knowledge does not increase output, if suddenly for some reason, demand for wooden shoes completely ceases. In that particular case, using one chisel and one unit of labor, 730 wooden shoes might be produced but due to a lack of demand this production does not actually take place and by consequence the extra knowledge has no value (nor has the previously existing knowledge which is not used productively either). Stated differently, knowledge might become less valuable over time, it might depreciate or, even better, it can get obsolete, when the goods in the production of which it was used are no longer in demand. This might have happened for instance with the knowledge associated with the creation of wooden shoes, while the knowledge contained in Newton's second law is still widely used today.

2. The second basic method of measuring knowledge can best be seen as the complete opposite of the first method. Rather than measuring knowledge by its effects on productivity, in this method knowledge is measured by determining the (value of the) resources that were used in its creation.[115] Equation (2.2) depicts a knowledge production

[114] The next section elaborates on issues regarding output and its measurement, which arise naturally in the study of growth.

[115] Alternatively, as what Mankiw (1997) in his well-known textbook refers to as one of the principles of economics, "The cost of something is what you give up to get it" (p. 5).

function in which new knowledge is created using a fraction (s) of the total labor force (L) and the existing stock of knowledge. Of course labor that is used in the production of new knowledge cannot be used to produce goods and the knowledge that is created in this way has been created at the expense of extra output. Hence, the value of some amount of knowledge equals the value of the production forgone. However, a problem similar to the one that has arisen earlier, arises with this approach to the measurement of knowledge as well. New knowledge increases future output and in this way, the new knowledge's discounted future benefits in the form of extra output may outweigh the temporary cost of lower production. This leaves the odd situation that the net costs of the creation of new knowledge are somehow effectively negative, which severely limits the usefulness of this approach to measuring knowledge. The idea that the value of knowledge is determined by what it costs to obtain it, remains appealing nonetheless. It probably does, because it is in line with the spirit of the R&D-based models of growth. In these models, maximizing economic subjects (e.g. firms) incur the cost of innovation in order to obtain a (temporary) monopoly and the corresponding profits. Hence, for a firm the value of the knowledge is either the discounted value of the stream of profits it generates or the cost of creating it. For society as a whole the value of course is greater but exactly how great cannot be easily ascertained, although a straightforward measure would be to account for the level of GDP and GDP growth by attributing these to their various sources. When capital, human capital, and the external productivity parameter, are accounted for, the remaining part must then be attributed to what here is defined as knowledge (see the previous paragraph; see also the work done on the elasticity of output with respect to R&D efforts discussed in Chapter 3 and other references in that chapter).

The above-discussed approaches to the measurement of knowledge have unsurprisingly confirmed the idea that measuring knowledge is not straightforward. As a consequence, there is, as of yet, no definitive answer regarding the way in which knowledge and the knowledge stock can/should be measured. This is not to say, however, that this discussion has been futile for it has provided an interesting lead to the pervading issue of scale effects in growth, namely the idea of knowledge becoming obsolete. How this plays a role in growth scale effects is discussed in the final section of this chapter.

5.2 Output, growth, the CPI and well-being

At first sight, the evidence on scale effects in growth presented in Chapter 2 is quite convincing for the idea that such an effect is absent. However, if one were to know that the data on GDP used in the various studies is seriously flawed, for example because the used GDP measures are unable to account accurately for the effects of quality improvements and the introduction of new products, one must be careful in drawing conclusions. Furthermore, these flaws warrant further attention for the measurement of variables that are important in the study of growth in general and regarding GDP in particular. The first paragraph (5.2.1) of this section subsequently is devoted to exactly that. Because in this discussion the issue of well-being is raised, the next paragraph (5.2.2) elaborates on the intricate relation between income and well-being, and the impact this has on the study of growth. Finally, as mentioned in the introduction to this chapter, Paragraph 5.2.3, points out the relevance of certain remarks made in the discussion in the previous paragraphs for the pervading issue of growth scale effects.

5.2.1 The CPI, output, and growth

There is - and has been - a lot of concern regarding the ability of the Consumer Price Index (CPI) to accurately measure the real change in the cost of living. In the U.S., this concern was undoubtedly an important reason for the origination of the "Advisory Commission to study the Consumer Price Index" or in short the CPI Commission. The CPI Commission has investigated the bias in the CPI and has concluded that it overstates the change in the actual cost of living by about 1.1 percentage points per year (Boskin, Dulberger, Gordon, Griliches and Jorgenson 1997, p. 78).[116] This inaccuracy has consequences for various economic issues such as the conduct of monetary policy, the cost of indexed spending and taxes, and the measurement of economic progress (ibid.). Regarding the measurement of economic progress, it is estimated that the upward bias in the CPI through mismeasurement of consumption (making up about two-thirds of GDP) causes real GDP to be understated by approximately a half percentage point per year (Boskin and Jorgenson 1997, p. 92).

[116] See also the final report from the 'Boskin Commission' to the Senate Finance Committee, "Toward A More Accurate Measure Of The Cost Of Living", available at http://www.ssa.gov/history/reports/boskinrpt.html (d.d. December 2003) (Boskin et al. 1996). In Boskin et al. (1997), Gordon and Griliches (1997) and Boskin and Jorgenson (1997), the members of the committee report on their findings. Finally, in Boskin, Dulberger, Gordon, Griliches and Jorgenson (1998) the members of the CPI Commission again discuss their findings and in addition rebut some of the critique that has been forwarded.

This statement requires some clarification of how the CPI is important in the measurement of economic growth. Therefore, the measurement of total output (GDP) and economic growth is elaborated on.[117] The basic principle behind measuring output and growth is quite simple. Assume that an economy or country produces only one good, apples. Last year, this economy was able to produce a total of ten million apples, while this year it has increased total production to eleven million apples. The country has then experienced ten percent economic growth, and if the population has remained constant, per-capita income has grown ten percent as well. Measuring GDP and economic growth for this hypothetical case is straightforward. Unfortunately, the real world is not as simple and clear cut as in this example. In the real world there are several factors complicating measurement. The most prominent of these factors is the fact that an economy's production is not homogenous at all but enormously diverse. That is, a country does not only produce apples but also for instance oranges. Determining economic growth if a country has increased its total production from 5 million apples and 10 million oranges, to 6 million apples and 11 million oranges is substantially less straightforward than in the example above and requires the ability to compare apples to oranges. To accomplish this feat, the value of oranges needs to be expressed in apples and vice versa. If the amount of different products produced in the economy is limited to two, this comparison can still be made rather straightforwardly. However, with an increase of the number of products (bananas, pears, cherries, peaches, strawberries, etc.) the matter gets incredibly more complicated. Therefore, in the actual measurement of total production and growth, a single unit of account is used, namely money (dollars, euros, yens, etc.). Unfortunately, using a monetary unit of account is not a panacea and is itself subjected to an important distortion. Just as the value of an apple expressed in oranges is not constant but varies over time, so too the value of money varies over time. How this effects the measurement of GDP and growth is explained using the equation of exchange well-known from monetary theory:

$$MV = PY \tag{5.7}$$

where M is the quantity of money, P is the price level, Y is real total output, and V is the velocity of money (by definition). The product of P and Y of course equals nominal GDP. Form the equation of exchange, it is clear that an increase in nominal GDP (P*Y) can have

[117] This issue is also discussed at some length in Pigou (1952, p. 50-55 especially).

two sources, an increase in the price level and an increase in real production.[118] This is even more apparent when the equation of exchange is rewritten to depict growth rates:

$$g_M + g_V = \pi + g_Y \tag{5.8}$$

where π is the rate of inflation (g_P). Equation (5.8) clearly shows that when the measured rate of inflation is biased upwards, the measured rate of economic growth (g_Y) is biased downward, and vice versa. The overstatement of the true increase in the cost of living by the CPI (as determined by the CPI commission) leads to an overstatement of the true overall rate of inflation (or the GDP deflator) and thus ultimately causes the measured level of total production and the measured rate of growth to be understated (cf. Boskin and Jorgenson).

The question now becomes why the CPI does not accurately describe the actual cost of living. As stated above, an important cause of error in the CPI is its inability to account properly for the effects of, quality change and the introduction of new products. This can be explained by first returning to the definition of a cost-of-living index: "a comparison in two time periods of the minimum expenditure required to achieve the same level of well-being." (Gordon and Griliches, p. 84). The trick now is to get the CPI (or any other measure of the cost of living) to be able to accurately determine how much a new product or a quality improvement in an existing product contributes to well-being and what the costs of this contribution are. This obviously is not straightforward. The CPI commission, however, proposes an interesting thought-experiment that sheds some light on the issue. The thought-experiment is to first determine how much well-being a certain level of income provides in 1970 with the 1970 basket and prices of goods, and to determine next how much more income would be required in 1995 with the 1995 basket and prices of goods, to provide an equal amount of well-being (Shapiro and Wilcox 1996, p. 117) (see also the Sears-catalog thought-experiment discussed in Gordon and Griliches). On the one hand, inflation has made money less valuable in 1995 than in 1970, and hence a higher income is required in 1995 to be able to obtain equal satisfaction. On the other hand, in 1970 modern products like CD players, Personal Computers, etc., were not available, so that in 1970 more income is required to obtain equal satisfaction. Overall, the CPI Commission conjectured that the percentage increase in income needed to be able to obtain an equal level of well-being in 1995 as in 1970, would be lower than the actual percentage increase in the CPI (Shapiro and Wilcox, p. 117). In the end, the

[118] If, as is usually assumed, the velocity of money (V) is constant, an increase in the quantity of money (M) accounts for the entire increase in nominal income.

CPI Commission estimated the bias resulting from quality change and new product introduction for 27 subcomponents of the CPI. They found a bias for example for Fresh fruits and vegetables of 0.60%, and for Alcoholic beverages of 0.15% (Gordon and Griliches, p. 86). Summing the weighted contributions of the bias in the 27 subcomponents, the CPI Commission found that the quality change new product bias in the CPI equals 0.60%, which is some 55% of the total bias in the CPI (1.1%). (Gordon and Griliches, p. 87).

Given the fact that the pervading issue of this thesis is scale effects in growth, the question that is really of interest, is how this bias affects the conclusion drawn from the discussion of empirical work in Chapter 2 that growth does not exhibit a scale effect. Although growth is understated due to this quality change new product bias in the CPI, this downward bias does not in itself mean that now growth does exhibit a scale effect. The most convincing piece of evidence in favor of the existence of a scale effect in growth is a growth rate that increases (or even accelerates) over time and the real (CPI bias adjusted) rate of growth only shows such behavior over time if the upward bias in the CPI showed an increase over time (see the empirical evidence in Jones 1995a).

For several of the 27 subcomponents of the CPI, the Commission has estimated the quality change new product bias over different time intervals. One method to do so, and indeed applied by the Commission[119], is to use so-called hedonic pricing, which is a useful instrument to examine the quality change bias (see for instance Griliches 1961a, Griliches 1971a and other studies in Griliches 1971b). A hedonic price index is a price index that makes use of hedonic functions (Triplett 1987, p. 632), i.e. relations between prices of different goods/services and the quantity of characteristics embedded in these goods/services (Triplett 1987, p. 630). In hedonic pricing, the price of a certain product is constructed from the price of a reference product, to which the prices of different quality characteristics are added or subtracted, where the price of each of these qualities is determined using multiple regression analysis. The price of a car is for instance determined by the amount of brake horsepower (BHP) it has, its (shipping) weight, its length, whether it has power steering and so forth (cf. Griliches 1961). When subsequently, over a certain period, the (average) price of cars has increased, this increase can be attributed to an actual increase in price, which is what the CPI should measure, but also to cars on average being equipped with more quality characteristics, which can be controlled for through hedonic pricing. That is, the CPI is obviously biased

[119] See Boskin et al. (1996), the final report from the 'Boskin Commission' to the Senate Finance Committee, "Toward A More Accurate Measure Of The Cost Of Living".

upward when the price increasing effect of extra quality characteristics are not accounted for accurately.

As stated, the CPI Commission has used the hedonic pricing method (and others) to estimate the quality change new product bias over different time intervals. The results of these estimations are shown in Table 7.

Table 7: Quality-change new-product bias for selected CPI (sub)components over time.[120]

CPI (sub)component	Estimated annual bias	Time interval	Relative importance (%)
Appliances, including electronic (Housing)	**3.6**	**1973-1994**	0.806
	5.6	**1994-1996**	
(Apparel and upkeep)	**-0.95**	**1965-1985**	5.516
	1.00	**1985-1996**	
New vehicles (Transportation)	**0.00**	**1970-1983**	5.027
	0.59	**1983-1996**	
Used cars (Transportation)	2.44	1967-1987	1.342
	0.59	1987-1996	
Motor fuel (Transportation)	**0.0**	**1974-1984**	2.908
	0.25	**1984-1996**	
Public transportation (Transportation)	**2.66**	**1972-1977**	1.523
	4.60	**1977-1982**	
	0.00	1982-1996	
Prescription drugs (Medical care)	3.00	1970-1995	0.891
	2.00	1995-1996	

Remarks: Major components in parenthesis (e.g. housing).

Table 7 shows that the bias in some components that weigh heavy in the CPI, e.g. Apparel and upkeep, New vehicles, and Motor fuel, has increased over time (**bold**). Hence, it seems that the overall bias in the CPI increases over time although the CPI commission draws no conclusion regarding the behavior of the quality change new product bias over time.

As noted by Baily and Gordon (1988), however, an increase in the quality change new product bias over time is but one of two ways in which the downward measurement error in GDP and GDP growth might increase over time. The other way is that the share in total production of products, which are subjected to large understating measurement errors, increases over time. This effect of structural change is often used in studies that discuss the so-called 'productivity slowdown' (see for instance Baily and Gordon 1988 and Gordon 2000). Griliches (1994) goes one step further and divides sectors in reasonably measurable (e.g. wheat production) and the rest (e.g. lawyer services; unmeasurable) (Griliches 1994, p.

[120] Source: Gordon and Griliches (1997).

10).[121] He subsequently states that "In the early post-World War II period, the situation was not all that bad: about half of the overall economy was "measurable" in this sense. By 1990, however, the fraction of the economy for which the productivity numbers are half reasonable had fallen to below one-third." (ibid.). If unmeasurable effectively means that actual production in these sectors is understated, as is assumed here, the overall downward bias in measured GDP and GDP growth has increased over time[122]. Both this increasing share in total production of unmeasurable production and the aforementioned increase in downward bias resulting from quality change and the introduction of new products contribute to the productivity slowdown. Gordon and Bailey, however, find that despite removing sectors with a measurement problem the post-1973 slowdown remains clear.

Hence, the question that is really of interest here surfaces again: do the measurable sectors and a hypothetical economy comprised of only these sectors do show signs of a scale effect in growth (whilst the unmeasurable sectors do not)? Part of this question can be answered by looking at the results of the estimation of the linear regression equation ln(Output per worker) = $\beta_1 + \beta_2$*Year + ε, which are depicted in Table E.5 in Appendix E. This table shows that a linear time trend is a statistically significant estimate (the highest p-value equals 0.012) for the development of value added per worker for all sectors both unmeasurable and measurable for the United States for the period 1950-1990.

This is confirmed in Figure 15 which graphs the natural logarithm of Value added per worker for the measurable sectors (Agriculture (Ag), Mining (Mi), Manufacturing (Ma), Public utilities (PU), and Transport and communication (TaC)) and GDP per worker constructed for a hypothetical economy comprising of only these sectors for the U.S. over the period 1950-1990.

[121] In the terminology of Griliches (1994), the sectors Agriculture, Mining, Manufacturing and Transportation and utilities are measurable, whilst Construction, Wholesale trade, Retail trade, Finance, insurance, and real estate, Government and Other services are unmeasurable.

[122] Of course this only holds for the period in which the share of the unmeasurable sectors in total output has increased, e.g. between the early post-World War II years and 1990 (see also Appendix E for data on the changing sectoral structure of the U.S. economy.

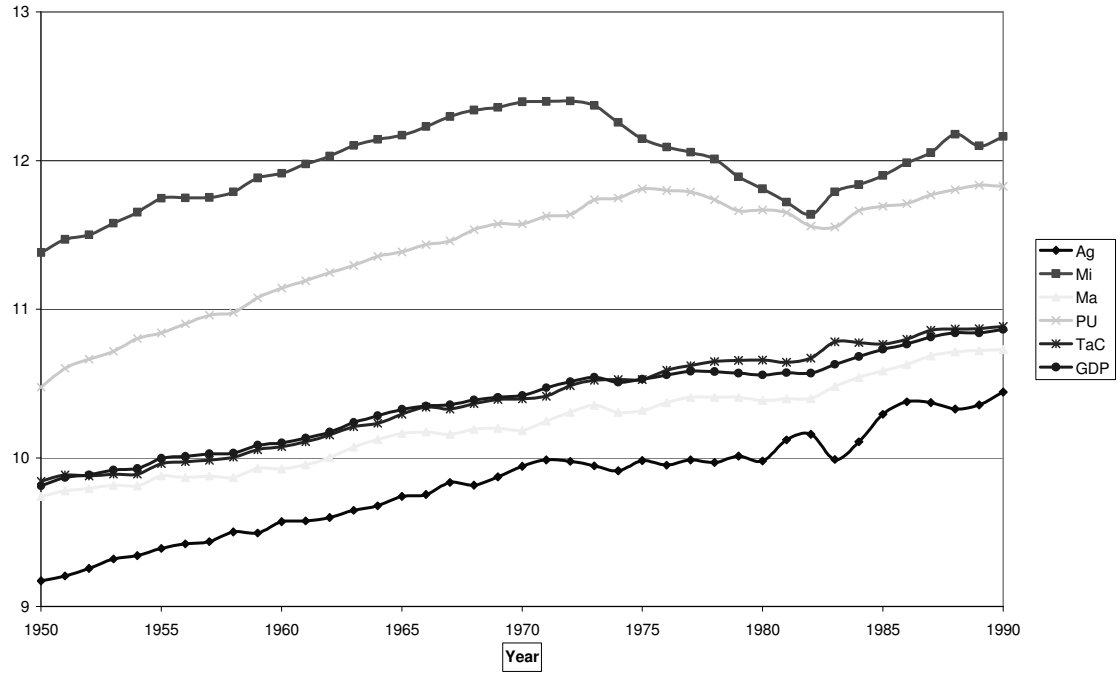


Figure 15, Value added per worker for the U.S. measurable sectors and a hypothetical economy.[123]

Figure 15 shows that sectoral value added per worker has followed a relatively steady path (although mining seems an exception and indeed, the linear estimation for mining is not significant at the one percent level (see also Table E.5 in Appendix E)). This is similar to the finding of Jones (1995a) that the growth rates of per-capita GDP of advanced OECD countries have been remarkable stable, especially since the 1950s, and do not show a persistent increase. Hence, the answer to the question whether the measurable sectors and a hypothetical economy comprised of only these sectors do show signs of a scale effect must be answered negatively.

To investigate further the possible effect of measurement problems on growth scale effects, the cross-sectional work of Backus, Kehoe and Kehoe is repeated for the measurable sectors. Figure 16 below, shows the growth rate of the Agricultural sector as a function of sector size for nine OECD countries (Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden, the U.K. and the U.S.).

[123] Source: van Ark (1996) and author's own calculations. See Table E.1 and E.2 in Appendix E.

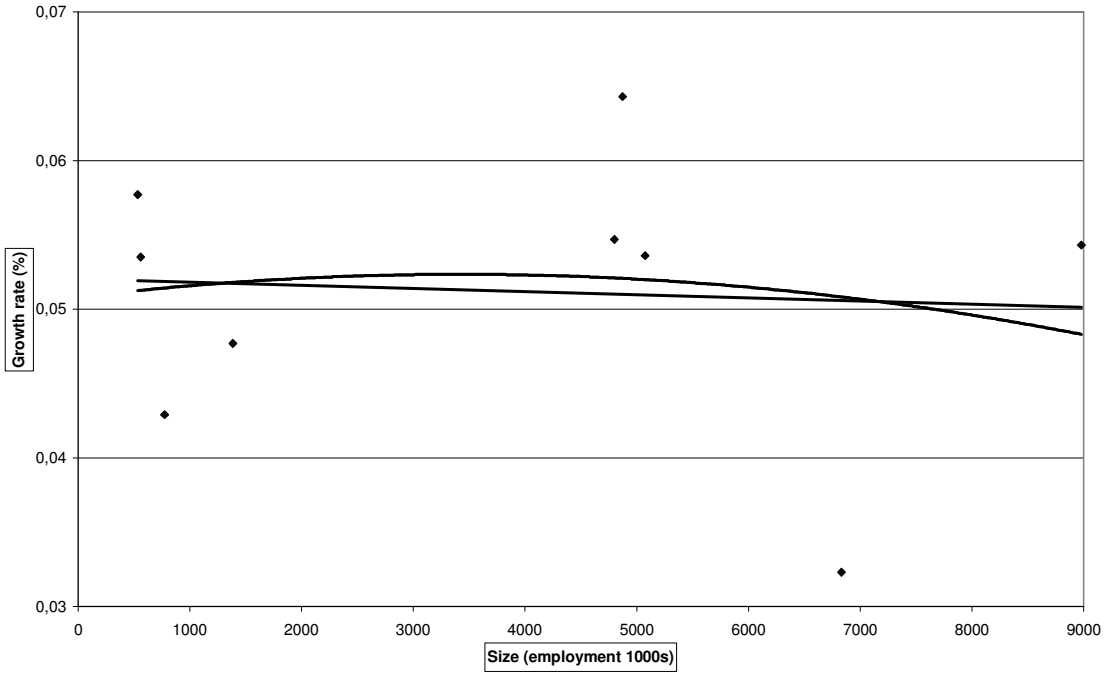


Figure 16, Sector size and sectoral growth rate for the Agricultural sector.[124]

It is clear from Figure 16 that the growth rate of the agricultural sector has no tendency to increase with sector size. The same goes for the other measurable sectors (Mining, Manufacturing, Public utilities, and Transport and communication) as figure H.1 in Appendix H shows. An interesting feature of Figure 16 and Figures H.1 is that the linear approximation reveals a tendency of larger sectors within a single type of sector (i.e. Agriculture, Mining, Manufacturing, Public Utilities or Transportation and Communication) to grow slower than smaller sectors of the same type. This is at odds with the prediction of the early models of R&D-based growth but also with the findings of Backus, Kehoe and Kehoe. Besides linear model, Figure 16 and figure H.1 also depict the quadratic approximation of the relation between sector size and the rate of growth of labor productivity. The differing results for this quadratic model, i.e. both hump-shaped and U-shaped, contribute little to Papageorgiou's finding of a hump-shaped relation between growth in GDP per capita and the size of an economy. Note, however, that generally neither the linear nor the quadratic model of the relation between sector size and labor productivity growth is significant in explaining observed variance. To conclude, a general model of productivity growth as a function of size is estimated, but again no significant relation could be established (see Table H.2).

[124] Source: van Ark (1996) and author's own calculations. See Table H.1 in Appendix H.

The overall conclusion regarding scale effects in growth is that the conclusion from Chapter 2, that growth at the aggregate level does not exhibit a scale effect can be maintained and is actually strengthened. That is, despite the fact that the CPI suffers from a quality change new product bias that increases over time, there is no evidence for scale effects in growth when attention is limited to sectors that produce goods which prices do not exhibit such a bias. Hence, one of the critical remarks regarding the conclusion that growth does not exhibit a scale effect, which was that measurement problems might underlie the observed absence of these effects, is rebutted here. However, more work along the lines of Jones (1995a) on scale effects at the sectoral level is warranted. This work should include long time-series data for many countries and for all sectors on for instance, investment (as a share of total sectoral output), average human capital per worker in the sector, and the number of people active in R&D in the sector. In addition, the work of Jones (1995a) should be repeated using aggregate data adjusted for the non-negligible and/or increasing quality change new product bias in some of the CPI's subcomponents.

5.2.2 Income, well-being, and growth

Having answered the question whether the original conclusion that growth does not exhibit a scale effect could be maintained, (unsurprisingly) yet another question emerges regarding the data used in the study of economic growth in general and to establish whether there is a scale effect in growth in particular. From the definition of a cost-of-living index given above[125] it follows that it is not really income (or better wealth) that should be of interest in the study of economic growth, but rather well-being (or happiness, utility, welfare, satisfaction, and so forth). When the CPI is constructed so as to more accurately account for the utility derived from quality changes in existing products and the introduction of new products, GDP and other output measures should be traded in for a measure like 'GDW (Gross Domestic Well-being)'. That is, the production of goods is ultimately aimed at satisfying needs and improving well-being and subsequently the real value of output should be measured by its ability to satisfy needs and improve well-being.

This in turn brings up the deeper question of whether GDP as it is currently determined is a good, or better-formulated, suitable, measure of well-being. This question is very close to what was assumed above to be one of the main reasons for the origination of the CPI

[125] That is, "a comparison in two time periods of the minimum expenditure required to achieve the same level of well-being." (Gordon and Griliches, p. 84).

Commission, namely concern regarding the ability of the CPI to accurately measure the real change in the cost of living.

Actually answering this question is beyond the scope of this thesis but several issues regarding the relation between income and well-being arise. Because some of these have an impact on the pervading issue of growth scale effects, they are elaborated on here[126]:

1. Besides income, consumer surplus, typically associated with an area under the demand curve, also plays an important role in well-being;
2. Well-being is not only a function of wealth and/or income but also of other factors like the environment, health, crime, etc. If the total production of an economy has increased at the expense of for instance a severely polluted environment, it is doubtful whether overall well-being has increased. Similarly, a higher GDP resulting from spending on for instance warfare, cleaning environmental pollution, etc., is unlikely to increase total well-being. Work on the CPI generally addresses these issues by pointing out that quantitative measures on for instance air and water quality, and crime rates have shown improvements over the last few decades (Gordon and Griliches, p. 88). However, rather than the actual state of the environment and the level of crime, a more important determinant of well-being is how people perceive these, e.g. the general feeling of unsafety is more important than the actual number of crimes committed. In addition, increasing income allows for the satisfaction of needs that are ever higher in the hierarchy. In fact, a good that once was a luxury over time and with increasing income turns into a necessity. This is what happened with for instance refrigerators, mobile phones and magnetrons, and might very well occur for such 'goods' as the environment, health, crime, etc.;
3. The exact relation between wealth and well-being is unclear. One would for instance like to know, whether well-being doubles with wealth, and whether the effect of an increase in wealth varies with the amount of wealth already obtained. Intuition tells us that a person's well-being increases more when this person's total wealth increases from 10,000 dollar to 20,000 dollar than when it increases from 100,000,000 dollar to 100,010,000 dollar or even from 100,000,000 dollar to 200,000,000 dollar;

[126] These issues are well established in (theoretical) literature but because of the limited scope of this chapter in general and this paragraph in particular they are discussed here without much reference to this (theoretical) literature. Nonetheless, some interesting references regarding especially the social context of consumption (so-called interdependencies) that deserve mentioning are, Veblen (1899), Duesenberry (1949), Leibenstein (1950) and Hirsch (1976), while Pigou (1952) has a broader perspective.

4. It is still difficult to determine the contribution of quality changes and the introduction of new goods to well-being. Again relying on intuition it is conjectured that the well-being product pioneers derived from owning the very first record player is almost equal to the well-being other (or perhaps the same) product pioneers derived from owning the very first CD player;
5. Closely related to this third issue is the fact that income's effect on well-being works partly in a relative way. That is, not only a person's absolute amount of wealth determines his well-being, but also his wealth relative to that of others (e.g. friends, family, colleagues, etc.). A person's well-being will be higher when he has a total wealth of 20,000 dollar and others are less wealthy (e.g. their total wealth equals 10,000 dollar), than when he has a total wealth of 20,000 dollar and others are wealthier (e.g. their total wealth equals 100,000 dollar);
6. The notion that a person would be happier owning an old-fashioned record player knowing that no better sounding and more user-friendly CD players are available, than he would be knowing that such wonders of modern technique, CD players, do exist.

Besides in theoretical work these six issues have also been discussed extensively in empirical work. From this work two important findings, which are most important to the study of growth in general and the issue of growth scale effects in particular, stand out:

1. for well-being absolute income matters less than relative income (see for example Easterlin 1974 and McBride 2001),
2. in periods with increasing income, an increase in well-being was lacking (see for example Easterlin 1974 and Blanchflower and Oswald 2003).[127]

In this context Abramovitz (1979) speaks of the 'Easterlin paradox', which is that "If one considers people in a given country, one finds what an economist would expect, a strong, consistent, positive association between income and happiness. […] On the other hand, if one compares countries with widely different income levels at about the same time, differences in reported happiness are small." (p. 7). The reasons underlying this paradox are numerous but often mentioned is that aspirations are relative and that well-being is subjective (subjective well-being) (cf. Easterlin 1974, Abramovitz 1979, van de Stadt, Kapteyn and van de Geer 1985, and Easterlin 1995). The implications of these findings go way beyond this paradox however. Layard (1980) for instance sees a negative externality of consumption (of a

[127] Other studies report similar findings as reported here, e.g. Layard (1980), van de Stadt, Kapteyn and van de Geer (1985), Schor (1991) Easterlin (1995), Oswald (1997), Frank (1999), and Easterlin (2001).

particular good) by one person reducing the level of well-being of other persons; "… many of the strivings of modern man (to educate himself, to migrate, to entertain well and so on) are in part self-defeating in terms of collective satisfaction […]" (p. 741). In much the same fashion, Easterlin (1974), Hirsch (1976), Scitovsky (1976), and Easterlin (1995) conclude that the case for growth as a way of increasing well-being is no longer so clear; "increasing the incomes of all does not raise the happiness of all." (Easterlin 1995, p. 44). Finally, Epicurus (± 300 B.C.) already concluded the following "If you wish to make Pythocles rich, do not give him more money, but diminish his desire." (Epicurus to Idomeneus, as cited on the first page of Zolotas 1981) and a daunting conclusion may be that human beings simply require 'more' just to keep the same level of well-being.

This disenchantment of growth (Abramovitz 1979) in turn has consequences for measurement issues in the study of growth in general and hence for the issue of growth without scale effects as well.

If well-being has been constant over say the last few decades while during this time total income has increased drastically, according to the definition of a cost-of-living index, there has been no growth, just an increase price in the overall price level. After all, during these decades the level of well-being has remained the same and hence the larger current income buys exactly the same level of well-being as did the lower income a few decades ago. Apparently, all that has happened over these decades is that, according to the definition of a cost-of-living index, the expenditure needed to maintain the level of well-being has increased. Taking the argument one step further, would be to conclude that all attempts aimed at accurately accounting for the bias in the CPI resulting from quality change and new product introduction are futile and the only way in which the CPI is able to accurately measure the real change in the cost of living is to simply use differences in the level of GDP over time (at least for the set of (advanced) countries for which the level of well-being has been relatively constant over the years).

This, however, is a very bold statement and rather one should argue that there are two extreme measures of variables like output (production) and derived variables like growth:

1. Happiness. Happiness is seen as a proxy (the best available) of well-being and thus anyone who is interested, in studying the (possible) increase of well-being over time, which often is the goal of studying economic growth, should look at data provided by for instance surveys of happiness;

2. Absolute (physical) output. Absolute output is a physical measure and should be seen as output of a well-defined product or class of products like wooden shoes.

Both these measures have weaknesses however; even though for many countries the overall level of well-being has not increased over the last few decades there has been technological progress and these countries have actually become better at producing countless products, and comparing productivity differences over time requires accounting accurately for changes in inputs and also for product changes like quality. The CPI subsequently should be seen as a measure of output (and growth) that combines the best of both these extreme measures.
The only problem that then remains for the CPI is whether it is able to keep combining the best of these two extreme output measures or not. The concerns of the CPI Commission with the ability of the CPI to accurately account for biases resulting from quality change and new product introduction and measure the real change in the cost of living should be seen in this light and seem to indicate that they feel the CPI has in fact not been able to do so.

5.3 Discussion and conclusions

The measurement issues discussed in this section have not been discussed just for their own sake, but also for the possibility that they shed light on important aspects of economic reality that may help the understanding - and modelling - of growth without scale effects.
The most obvious result of the discussion in the two previous sections for the pervading issue of growth scale effects is that the conclusion, drawn from the empirical work on these effects discussed in Chapter 2, that growth does not exhibit a scale effect stands even more firmly than before. That is, the quality change new product bias which causes the CPI to overstate the true cost-of-living does not lead to a new conclusion regarding the existence of growth scale effects and this is despite the fact that this bias has increased over time (see Paragraph 5.2.1).
The discussion in the previous two sections has raised several other issues however, which could and quite possibly even should be important for the study of growth without scale effects.

5.3.1 Knowledge, knowledge production, obsolescence and creative destruction

From the discussion in Section 5.1 two issues important for the study of growth scale effects emerge, (1) the definition of knowledge and how it combines with the external productivity parameter in the production function (e.g. multiplicative, (E*A) vs. additive, (E + A)), and (2) the idea that knowledge becomes obsolete because it is used for the production of goods which are no longer in demand. The most important implication these findings have for the study of simply regards the formulation of the knowledge production function. In the standard knowledge production function of the early models of R&D-based growth (e.g. Romer 1990, Grossman and Helpman 1991, and Aghion and Howitt 1992) new knowledge is created using a fraction (s) of the total labor force (L) and the existing stock of knowledge (see equation 2.2):

$$\dot{A} = \delta sLA \tag{5.9}$$

where δ is a parameter determining the arrival rate of new knowledge. In this formulation of the knowledge production function there is a scale effect in growth because the production of knowledge is a function of the existing stock of knowledge and there is a strong intertemporal knowledge spillover in the R&D process.

This formulation changes, however, when the findings regarding knowledge's definition (Paragraph 5.1.1) and its obsolescence (Paragraph 5.1.2), are taken into account. A straightforward adaptation of the standard knowledge production function subsequently would be:

$$\dot{A} = \delta sL(A + E) - dA \tag{5.10}$$

In this equation productivity in the R&D process depends not only on the existing knowledge stock but also on the external productivity parameter and knowledge is accumulated in much the same way as capital is accumulated in the standard neoclassical model of growth (e.g. Solow 1956), i.e. new knowledge is added by using part of the available productive resources creating it while old knowledge depreciates, i.e. becomes obsolete, at rate d. Consequently, the prediction of a scale effect in growth that was so apparent in the knowledge production function of the early models of R&D-based growth is less clear. Indeed, the exact specification of the knowledge production function may be such that, as observed, growth does not exhibit a scale effect. (Equation 5.10 is just a highly stylized example, which,

assuming that d and E are exogenous parameters, unambiguously predicts a scale effect in growth.)

Besides leading to the straightforward reformulation of the knowledge production function depicted above, the idea that (1) it is unclear how (R&D-based) knowledge (A) combines with the external productivity parameter (E), e.g. multiplicative (see equation 5.5) or additive (see equation 5.6 and also equation 5.10), and the idea that (2) knowledge somehow becomes obsolete, raise some additional issues that deserve discussion.

1. Regarding the first idea, it was already mentioned in Paragraph 5.1.1 that the effect an R&D-induced increase of the knowledge stock (A) has on total factor productivity (growth) depends on the exact formulation of TFP as function of E and A (i.e. TFP(E,A)) and that this could play a role in the reconciliation of the findings of Porter and Stern (2000) with those of Jones (1995a) (see Chapter 3; see also Lach 1995, Verspagen 1996, and Bloom and van Reenen 2002). Indeed, a straightforward explanation for the apparent discrepancy between countries' ability to produce knowledge and their ability to translate this knowledge in productivity (growth) is that (R&D-based) knowledge and the external productivity parameter combine in an additive manner to form TFP. When E (the external productivity parameter) and A (knowledge) combine in this particular way to form TFP, any contribution of newly created knowledge to TFP may virtually vanish in the much larger contribution E makes to TFP: $E >> A$; $A/E \rightarrow 0$. That is, when other factors, e.g. political variables, civil liberties etc. (see Sala-i-Martin), play a much large role in TFP (growth) than knowledge (A) does, the limited elasticity of TFP with respect to knowledge (R&D, patents, and so forth) is not much of a puzzle.
2. The second idea, i.e. the idea that knowledge becomes obsolete because it is used in the production of goods which are no longer in demand, raises two interesting issues.
   (i) First of all, the relation between this idea of knowledge obsolescence and the concept of dilution of knowledge through technological distance. Generally, a certain piece of knowledge can be applied productively in the production of numerous goods. When in this situation demand for a product decreases, knowledge keeps (much of) its value because it can still be used productively in the production of other goods that do not face declining demand. Suppose, however, that knowledge is product-specific and is only used to its fullest (productive) potential when used in the production of the specific good for which it was originally conceived, and is much less useful, i.e. diluted, when used in the production of other products. Then, a large decline in demand for a specific

product has the effect of causing the concomitant product-specific knowledge to lose much of its value (as measured by its effect on output (per unit of productive resource)). Hence, with dilution of knowledge through technological distance, (product-specific) knowledge is likely to become obsolete when the goods in the production of which it was used are no longer in demand.[128]

(ii) Second, the concept of knowledge becoming obsolete as a result of demand factors, leads to the logical extension of the rate of obsolescence being determined endogenously: endogenous (R&D-based) technological progress, i.e. knowledge creation, leads to, amongst others, the introduction of new goods and demand shifts (see Chapter 4) causing knowledge to become obsolete (creative destruction), and the faster new knowledge is created, the faster also existing (product-specific) knowledge loses its value; $d = f(\dot{A})$, with $\partial d/\partial \dot{A} > 0$.[129]
The idea of an endogenous rate of knowledge obsolescence in turn raises yet other interesting issues.

1. A first regards the pervading issue of growth without scale effects, and is simply that it is much more likely that growth does not exhibit a scale effects when there is creative destruction, i.e. when the rate of obsolescence of knowledge is a function of the rate at which new knowledge is created (see also equation 5.10).

[128] Note that the combination of demand and dilution is only one way in which (product-specific) knowledge can become obsolete and that there are other sources of obsolescence. Take for instance the important idea behind R&D-based models of growth that firms conduct R&D in an attempt to achieve market power and make a (monopoly) profit. When the knowledge granting a firm its market power, diffuses to their competitors, the value of this knowledge for the firm, as measured by the monopoly profits, of course is diminished / suffers from obsolescence. Further note that knowledge obsolescence through demand and dilution not only expresses itself in knowledge losing its usefulness in the production of actual goods, but also in the creation of new knowledge, i..e. it also has an effect on intertemporal psillovers (cf. Peretto and Smulders). The idea is that knowledge asssociated with products for which is low, e.g. wooden shoes, is most useful in the creation of knowledge regarding this type of product and much less helpful in the creation of knowledge regarding other products, e.g. DVD recoreds, LCD screens and so forth, for which demand is not low (and actually is expected to increase. The idea of obsolescence of knowledge through dilution (with an important role for technological distance) is thus closely related to the Peretto-Smulders model and the idea of endogenous network particpation discussed in Chapter 3.

[129] The idea of an endogenous rate of knowledge obsolescence is supported in empirical work.
Goto and Suzuki (1989) for instance find that the rate of obsolescence of R&D capital (i.e. the knowledge stock) as measured by the inverse of the life span of patents (see also Bosworth 1978) differs over industries. In industries in which technology advances rapidly, e.g. precision machinery and communications equipment and related products, the rate of obsolescence is much higher than in less high-tech industries like food and stone, clay and glass. They actually find that in industries with high R&D intensity, the R&D stock is small (Goto and Suzuki, p. 557).
Caballero and Jaffe (1993) report similar results and they have applied their findings to a model in which creative destruction, i.e. the endogeneity of the rate of knowledge obsolesce, plays an important role.

2. A second implication regards the weak link between knowledge and knowledge production on the one hand and the level of productivity and productivity growth on the other. Studies like that of Porter and Stern (2000) and other studies discussed in Chapter 3 (e.g. Bloom and van Reenen 2002) have used knowledge stocks corrected for obsolescence to investigate the relation between knowledge (measured using (past) R&D expenditures and (past) patents) and TFP. They have not, however, taken into account the fact that the rate of obsolescence of knowledge is probably related to the rate at which new knowledge is created, i.e. that the rate of obsolescence of knowledge is endogenous. That the endogeneity of knowledge obsolescence, i.e. creative destruction, can nonetheless be important is evidenced by the study of Caballero and Jaffe (1993) already mentioned in footnote 129. In their model, the rate of knowledge obsolescence is endogenous and when estimated empirically is able to describe closely actually observed rates of consumption/productivity growth and knowledge growth (measured using patents). Hence, future work that investigates the relation between knowledge and productivity (e.g. the patent-productivity relation; see for instance Lach 1995 and Verspagen 1996) should take the endogeneity of the rate of obsolescence of knowledge into account.

The conclusion that emerges from the several implications and issues raised and discussed here is simply that still little is known on the role (R&D-based) knowledge plays in TFP and that future research along the lines of for instance Bloom and van Reenen is warranted.

5.3.2 The measurement of output and growth scale effects

One of the insights of Section 5.2 is that there are different approaches to the measurement of output (income), e.g. happiness (well-being), monetary units (GDP as it is currently measured, i.e. using the CPI), and absolute (physical) output. This has far-reaching implications for the study of growth scale effects and the study of growth in general because each of these three approaches tells a different story about growth:

- Happiness. For several countries well-being as measured by surveys of happiness has not increased the last few decades so that in this extreme approach growth has been absent;
- Monetary units. In the current and widely used approach of measuring output as GDP (a monetary unit) corrected for price-level changes using the CPI growth has been persistent;

- Absolute physical output. This other extreme approach suffers from a lack of data so no conclusion regarding the observed growth experience can be made using this approach.[130]

The obvious implication here is that the way to study growth, e.g. which measures to use, is dependent on how one views the process of growth; is the goal of growth to increase well-being, is it to maintain a certain level of well-being, or is it to improve the score on a set of (fundamentally subjective) criteria like civil liberties, infant mortality and income inequality? Dependent on the answer to this and analogous questions, a choice can be made to focus attention on happiness, absolute output or a combination of these two, e.g. a monetary unit (GDP), and try to explain their behavior over time using the familiar methods, e.g. creating theoretical models and test them with empirical work.

For the issue of growth scale effects, the implications of the three approaches to the measurement of output/income are even more far-reaching. Take simply the conclusion regarding the existence of these effects. This conclusion may vary with the measure used just like the growth experience differs with the measure used (happiness vs. GDP): since the level of happiness has not grown at all over the last few decades, growth does not exhibit a scale effect when either happiness (well-being) or, as concluded in Chapter 2 and in Section 5.1, GDP (a monetary unit) are used as the respective measures of output and growth. Growth however, may exhibit a scale effect when physical (absolute) output is judged to be the relevant measure. That is, just like the level of income has increased when the measure used is GDP but not when happiness is the measure used, growth may exhibit a scale effect when absolute output is the measure used but not when GDP is the measure used. What is more, there is no a priori reason to assume that knowledge's nonrivalrous nature should reveal itself in a monetary output measure. Indeed, happiness and physical (absolute) output seem to be no less qualified for this task and there may even be a case to argue that physical output being the purest of measures is in fact the most qualified.

As simple as this last notion is, it indeed has a far-reaching implication, for in it may lie a reason why, when GDP is the measure used, scale effects are absent at the aggregate level. The idea is simply that growth does exhibit a scale effect when measured using physical units but not when measured using monetary units (GDP) or happiness (well-being) for that matter.

[130] The only data that comes close to data concurring with the idea of absolute output (technical efficiency) is data that has been used in establishing the effects of learning by doing and experience curves. Especially interesting in this respect is the building of the Liberty Ships for World War II in which for a prolonged period of time ships were build using the same blueprints (and of course labor and capital) (cf. Searle 1945, Rapping 1965 and Lucas 1993, and Thompson 2001a for a more critical view). However, even with this pure data difficulties arise because data on R&D inputs are lacking and it is difficult to distinguish learning-by-doing effects from the effects of (process) innovation (R&D effort).

This in turn has certain similarities with the mechanism by which, in the simple and highly stylized model presented in the previous chapter, growth at the aggregate level does not exhibit a scale effect, while sectoral growth does (or at least can). What is more, the discussion in Section 5.2 has made it more plausible that there indeed is an intricate relation between factors like consumer preferences, income, relative prices and technological progress. The conclusion is then that this intricate relation partly expresses itself in the (relative) prices of products and thus monetary measures of output and in this way may cause growth scale effects to be absent (when a monetary measure like GDP is used but not when the pure measure of physical output is used).

5.3.3 Well-being over time and creative destruction

Much of the 'problem' with the monetary measurement of output as opposed to the physical measurement is the result of consumer preferences or taste; as Pigou (1952) notes, "The dividend is an objective thing, consisting in any period of such and such a collection of goods and services that flow into being during the period. Since it is an objective thing, we should naturally wish, if we were able, to define changes in the size of it by reference to some objective physical unit, and without any regard to people's attitude of mind towards the several items contained in it. I do not mean that changes in public tastes would be thought of as incapable of affecting it by causing changes in the objective constituents of the dividend. I mean that, *given those objective constituents*, the size of the dividend should depend on them alone, and not at all on the state of people tastes." (p. 50, italics original).
How taste or consumer preferences can be important is shown in this paragraph where it is used to account for the constant level of well-being observed over the last few decades and related to two (additional) manifestations of creative destruction (see also Paragraph 5.3.1).

1. As noted in Paragraph 5.2.2, income and consumption cannot be viewed separately from their social context (there are interdependencies or preferences are interdependent). Layard even goes so far as to see a negative externality associated with consumption: consumption (of a particular good) by one person reduces the level of well-being of other persons. That is, people suffer from the consumption of others even though their income and other factors that determine their well-being have not changed. This in turn can be related to creative destruction (par excellence) in the following way: the introduction of new products diminishes the amount of well-being people derive from a pre-specified consumption bundle or stated differently the introduction of new goods brings about a

change in consumer preferences (tastes). As noted in Paragraph 5.2.2, a person would be happier owning an old-fashioned record player knowing that no better sounding and more user-friendly CD players are available, than he would be knowing that such wonders of modern technique, CD players, do exist. Creative destruction thus not only encompasses the replacement of older products by new products, which are better able to satisfy the needs of consumers than the older products, but also the partial destruction of the capability of these older products to supply its consumers with well-being in turn the result of the changing preferences. Viewing creative destruction in this way provides a simple explanation for the finding that over the last decades well-being has not increased (while total output, income and consumption have): over these decades the continuous introduction of new products and the actual consumption of these products by some people has offset any increase in overall well-being resulting from the increase in income occurring simultaneously. That is to say, preferences of consumers (consumer taste) have reacted to new goods by diminishing the positive effect on well-being of previously consumed goods. This view is of creative destruction is of course seriously at odds with the Dixit-Stiglitz model in which consumers have a preference for variety and in which consequently well-being is increased when more options are available. Empirical evidence, nonetheless, provides support for this last idea (and other ideas formulated above). For instance, Falkinger and Zweimüller (1996) find that the number of products consumed increases with per-capita income (see also Falkinger and Zweimüller 1997), and Bils and Klenow (2001) find that "… variety has increased by perhaps 1 percent per year over the past 40 years [but] that most of this growth occurs in just the past 20 years […]" (p. 274).

2. Also noted in Paragraph 5.2.2 is that income is not the only determinant of well-being but consumer surplus typically associated with an area under the demand curve is important as well. The reason why there is a consumer surplus is that consumers have different preferences (tastes) and that producers are unable to (i) let each consumer pay for their products the exact amount corresponding to its preferences, and (ii) tailor their products to the exact preferences of each consumer. Rather producers sell products with general features for a general price so that some consumers pay a price for a certain product which does not fully reflect the well-being they derive from consuming this product; there is consumer surplus. If, in contrast, producers are able to discriminate perfectly between consumers by charging different consumers different prices for the same product, by supplying different consumers with (slightly) different products (of which the price of

course also differs) or use a combination of these, they are able to effectively eliminate consumer surplus. For so far as this consumer surplus is not fully converted to profits (producer surplus) overall well-being may suffer. This in turn may actually have happened. Many of the new products introduced over the last decade may have been products that narrowed the gap between widely differing consumer preferences on the one hand and the limited number of products available to satisfy a certain need. That is, where previously a single product was meant to serve a market for a given need of size 10,000 (people/dollars/etc.), now the average market size for a given need served by a single product may have declined to 5,000. The result of the introduction of new products may thus ultimately be that total output (demand) increases by exactly the same amount by which consumer surplus has decreased leaving the overall amount well-being unaffected. An interpretation of creative destruction related to the picture sketched here is that the creative process by which new goods are introduced, also destroys consumer surplus.

5.3.4 Concluding remarks

This chapter has sketched a the contours of depiction of reality in which the absence of growth scale effects, where logically such effects are expected as result of knowledge's nonrivalrousness seems less of a puzzle than before, e.g. at the beginning of this thesis. This is because the chapter has pointed out several factors that may underlie the absence of such effects. Most interesting however, is that all these factors are somehow related and that what can be broadly be referred to as well-being is what binds them.

For instance, the idea of an intricate relation between consumer preferences, income, relative prices, technological progress (in a broad sense) etc., is supported in the findings of this chapter. This in turn lends credibility to the model developed in the previous chapter in which this intricate relation causes the absence of scale effects in growth at the aggregate level. Similarly, human nature (taste, individual's welfare or well-being function, and so forth) seems to drive a wig between output measured in absolute physical units and output measured using a monetary measure like GDP, which is that any measure other than the pure physical one entails some sort of value judgement in turn determined by taste (consumer preferences) (cf. Pigou).

The main conclusion following this chapter is consequently simply that certain properties of peoples' well-being functions (or, more general, human nature), e.g. the nonhomotheticity of preferences, the effect of the social context in which consumption takes place etc., with all the

associated intricacies are an important factor underlying the observed absence of growth scale effects. The exact role peoples' well-being functions play in the complete mechanism by which this feat is achieved is not yet clear, although the previous chapter has shown one possible way. More broadly, the role of individuals' well-being functions in this mechanism can best be described as giving rise to creative destruction, which in this mechanism manifests itself in three different forms:

1. (productivity) growth in combination with the introduction of new/improved products, i.e. technological progress in broad sense, leads to a changing composition of consumer demand giving rise to the familiar concept of the product life cycle in which demand for a product first takes off then stabilizes and ultimately declines asymptotically to its final value. This is the form of creative destruction familiar from Chapter 4 and Paragraph 5.3.1;
2. The second form of creative destruction is that the introduction of new products actually diminish part of the ability to provide well-being of the already existing goods (see Paragraph 5.3.3);
3. The third and final form in which creative destruction might manifest itself and which is relevant for the issue of scale effects is that the introduction of new products increasingly enables firms to appropriate some of the surplus previously in the hands of consumers to some extent effectively destroying part of the consumer surplus, and thus, overall, diminish well-being.

Previously doubt may have existed whether the intricate relation between consumer preferences, income, relative prices, technological progress etc., i.e. the first form of creative destruction, was actually single-handedly powerful enough to cause scale effects to vanish at the aggregate level. Now, however, there seems to be a combination of three different manifestations of creative destruction working against growth scale effects. Stated differently, there are even more intricacies in the relation between consumer preferences, income, relative prices, technological progress etc., than was apparent from the discussion in Chapter 4. The premise regarding what causes the absence of growth scale effects subsequently is that there is an intricate relation between consumer preferences, income, relative prices, technological progress etc., which gives rise to various form of creative destruction, and ultimately causes growth scale effects to vanish (at the aggregate level).

The next chapter elaborates on this and tries to merge the findings of this - and especially the previous - chapter in a full theory of growth without scale effects.

6 The product life cycle, demand, technological progress, and growth without scale effects

The simple toy model of Chapter 4 has drawn attention to the product life cycle and the role it may play in the issue of growth without scale effects. Hence, the first section (6.1) of this chapter takes a closer look at this well-known phenomenon and relates different concepts of it to a single broad life cycle, what is here referred to as the 'standard life cycle'. The second section starts by discussing the idea of hierarchic demand and the relation between this concept and demand saturation and nonhomothetic preferences in Paragraph 6.2.1. In doing so, it extends the rather brief discussion of saturation and nonhomothetic preferences already presented in Chapter 4. In Paragraph 6.2.2 subsequently, studies that have used hierarchic demand and formalized ideas like the ones presented in this chapter and thesis and actually incorporated them in models of growth are discussed. The chapter ends with Section 6.3, which presents a discussion and some conclusions.

6.1 The product life cycle

In Chapter 4, demand for a product in the simple model was said to follow a life-cycle pattern. Within the context of that chapter is was intuitively clear what exactly constitutes a life cycle; over its lifetime, a product (or probably better demand for a product) goes through a cycle in which five consecutive phases can be discerned:

1. The development phase in which the product is introduced and demand for it grows slowly.
2. The growth phase during which demand for the product takes off and grows rapidly.
3. The maturity phase in which the growth of demand levels off and demand reaches its peak.
4. The decline phase in which, demand, just having reached its peak, starts to decline with increasing speed.

5. The obsolescence phase during which total demand declines slowly and asymptotically to its final level (e.g. zero).[131]

This intuitive life cycle is the one that is of interest in this chapter and this thesis and will be further referred to as the standard life cycle already mentioned above. In the literature, however, the concept of a life cycle has been applied in a much broader perspective than simply that related to the development of demand for a single product over time (see also footnote 131). First of all, (i) the life cycle has been (and can be) applied not only to individual products but to other levels of analysis as well, e.g. to product classes, to industries etc. Second, (ii) the life cycle has been placed in the context of (i) the diffusion of technology, (ii) the geography of production and trade, and (iii) the evolution of industries.
These three contexts are discussed in the next three paragraphs (6.1.1 - 6.1.3) and their link to the standard (intuitive) life cycle will be made apparent.

6.1.1 The diffusion of technology

Griliches (1957) studies the patterns of diffusion of a new technology, hybrid corn, across U.S. states and hence provides a useful starting point for this discussion of the life cycle in the context of diffusion of technology (see also Dixon 1980). He plotted the percentage of total corn acreage planted with hybrid seed as a function of time and found an S-shaped curve similar to the one Aoki and Yoshikawa use:

[131] To prevent beforehand possible confusion it must be noted that the number of (distinct) phases in the life cycle is not an absolute number. For example, the product life cycle is also often said to exist of three stages (or phases). The first stage is then said to be something like the take-off stage in which demand for the product takes off and grows almost exponentially (e.g. the take-off phase), the second stage is then the maturity stage (or something similar) in which total demand stabilizes, and the third and final stage in this three stage scenario is referred to as the decline stage in which demand declines asymptotically to zero.

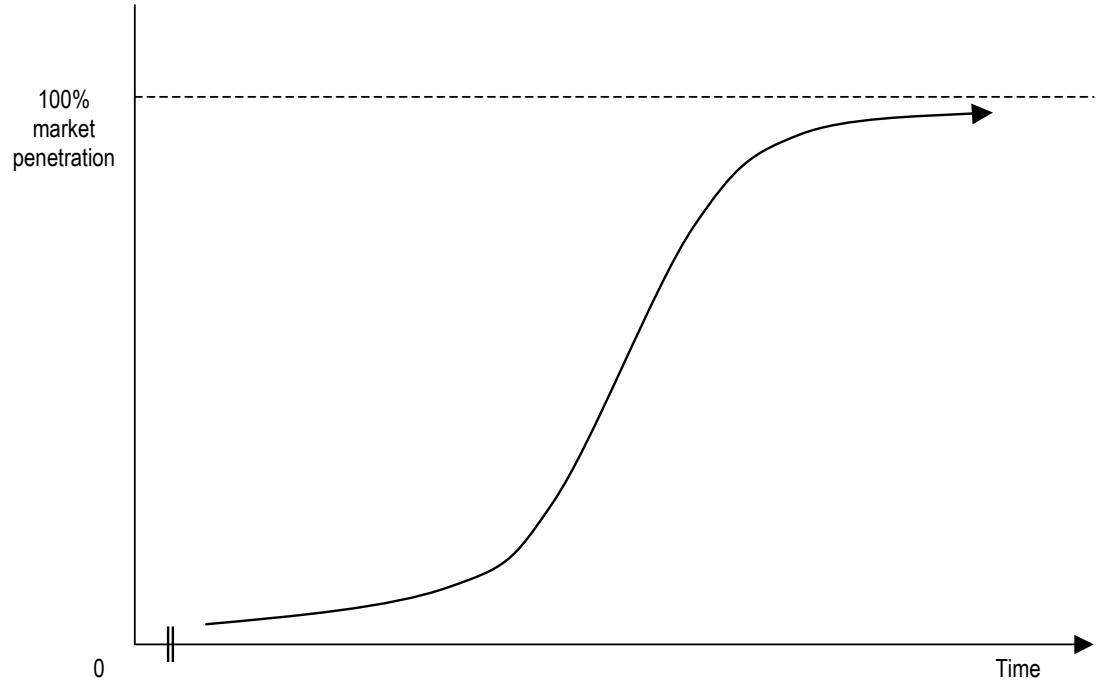


Figure 17, Percentage of use (market penetration) as a function of time.[132]

Griliches (1957) further found that a logistic curve accurately describes the observed pattern of the market penetration of the new type of corn seeds. That is, this S-shaped curve, which is typically associated with the life cycle of product, industries etc., can be described by the following equation:

$$P(t) = \frac{K}{1 + e^{-(a+bt)}} \tag{6.1}$$

where P is the market penetration rate (in Griliches' case the percentage planted with hybrid seed), K is the ceiling or equilibrium value, b is the rate of growth coefficient, and a the constant that positions the curve on the time scale, and t denotes time (Griliches 1957, p. 504). The logistic function has several interesting features the most prominent of which are that (i) it is asymptotic to both zero (0) and the ceiling (K) and that (ii) the rate of growth depends on both the distance to the ceiling and the level already achieved. This last feature can simply be shown by taking the time derivative of P(t) and with some rewriting the growth rate is given by:

$$\frac{dP(t)}{dt} = \dot{P}(t) = bP(t)\left(1 - \frac{P(t)}{K}\right) \tag{6.2}$$

[132] See Griliches (1957, p. 502). See also Figure 12.

When, as discussed by Pasinetti, market penetration of a product follows the logistic pattern described here and confirmed for many products in empirical work[133], total demand for this product follows the same pattern. Product demand than goes through three distinct phases, (1) the take-off phase in which demand growth rises, (2) the maturity phase in which the growth rate of demand stabilizes, and eventually (3) the stagnation phase in which the growth rate of demand declines asymptotically to zero. Hence, the link to a life cycle is clear.
What is more, the logistic pattern of diffusion can easily be extended to incorporate the standard life cycle discussed above. When the (market) diffusion of a new (or improved) product like hybrid corn follows a logistic pattern, logically the product it replaces, i.e. other types of corn (e.g. normal corn), follows the exact opposite pattern. In fact, the declining rate of percentage usage of the old product (P_O) as opposed to that of the new product corn (P_N) can probably be straightforwardly described by a straightforward adaptation of equation (6.1):

$$P_O(t) = K - P_N(t) = K - \frac{K}{1+e^{-(a+bt)}} \tag{6.3}$$

where b and a are the exact same coefficient as in equation (6.1). Figure 18 on the next page depicts the gradual replacement of the old product (normal corn) with the new product (hybrid corn). The idea is that the summation of both curves adds up to the ceiling (K). The process of replacement of old products of course is a continuing one and someday hybrid corn might be replaced by a new type of corn (e.g. genetically modified corn) or by another product like rice. The idea of diffusion (market penetration) and eventually replacement is what constitutes the standard life cycle familiar from Chapter 4 and discussed above as is shown in Figure 19 on the next page.

[133] The S-shaped pattern of diffusion indeed has been found for very many products; see for instance Aoki and Yoshikawa 2002, and Stoneman 2002 for an overview of some of this work).

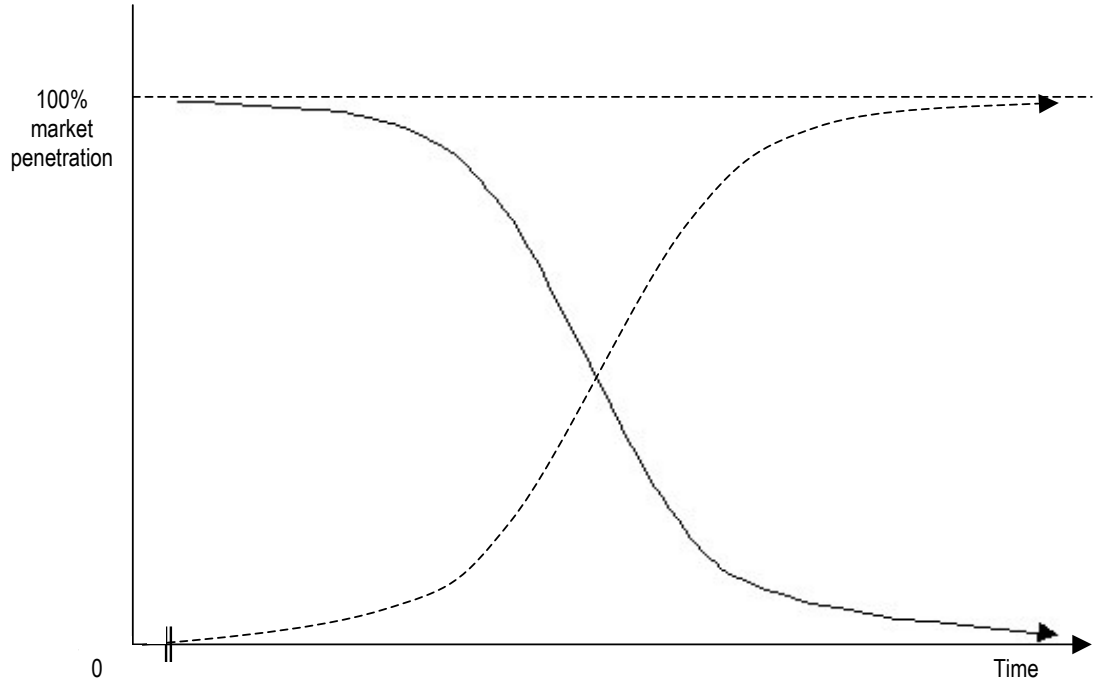


Figure 18, Diffusion of new - and replacement of old - products.

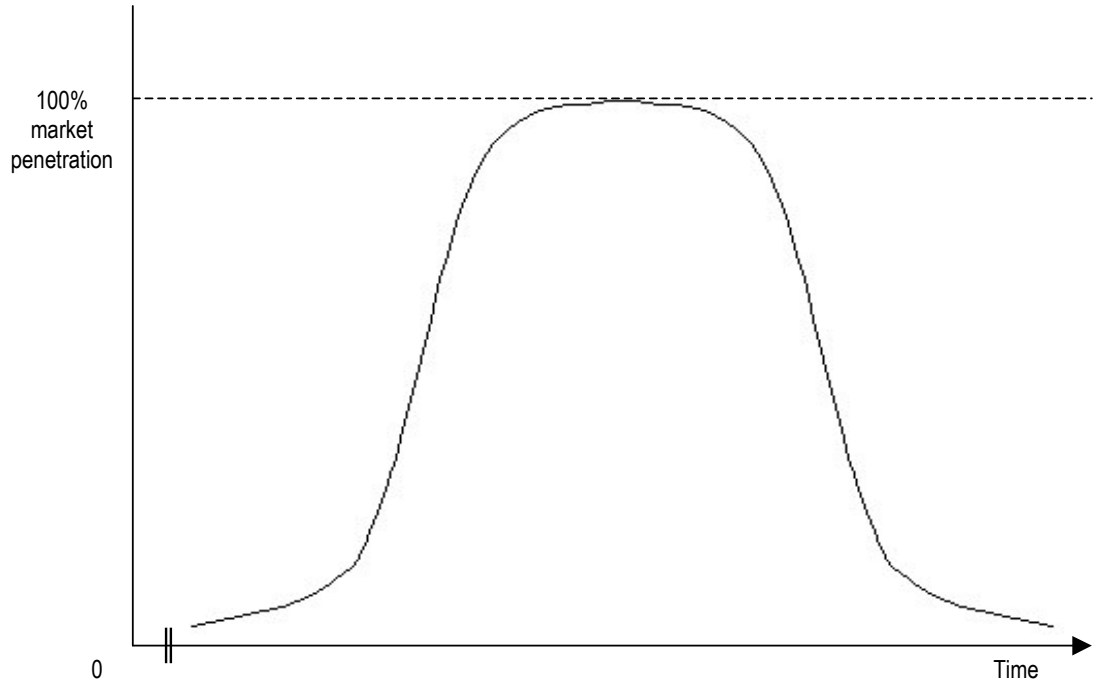


Figure 19, Market penetration over the life cycle.

In Figure 19, various phases can easily be discerned especially those associated with the take-off, stabilization and decline.

6.1.2 The geography of production and trade

Vernon (1966) ties the life cycle concept to the location of production and the (resulting) patterns of trade (see also Hirsch 1967, Wells 1972a and other studies in Wells 1972b). The idea is that over time, as a product matures, production moves from its location of origin (i.e. the United States) to other places and that patterns of trade change correspondingly. As this process continues, a geographical life cycle emerges. The result of this geographical interpretation of the life cycle concept is a pattern as depicted in Figure 20.

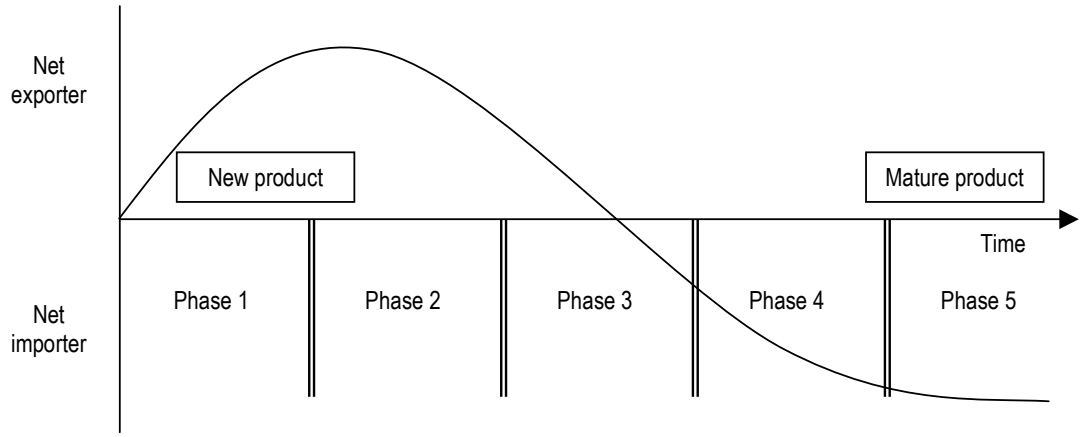


Figure 20, The geographical life cycle for U.S. production and trade.[134]

Starting with its birth and until it reaches maturity, the location of production and the pattern of trade of a product moves through five distinct phases. Starting point for Vernon's interpretation of the life cycle concept was the observation that many of the products newly introduced in the twentieth century have their origin in the United States (e.g. semiconductors, (personal) computers, etc.). The reason underlying this observation is, according to Vernon, the fact that the U.S. provides a large market for new products due to its large population and the high per-capita income (see also Chapter 2; cf. Sokoloff and Ades and Glaeser) and that domestic (U.S.) producers have superior understanding of the local market. Then, after the product has been conceived, the cycle starts:[135]

1. In phase 1, all production takes place in the U.S. and the U.S. serves any overseas demand (e.g. from Europe[136]) through exports. This situation does not last indefinitely, however,

[134] figure adapted from Wells (1972a, p. 15).
[135] This discussion of the geographical life cycle and its five phases is based on Wells (1972a, p. 11-15).
[136] Instead of Europe, Japan could just as easily have been used as an example. However, in the 1960s, the time when Vernon's theory was developed, the level of development of Japan was still lagging far behind both that of the U.S. and Europe, and hence Vernon's theory is focussed on the U.S. on the one hand and Europe - and not Europe and Japan - on the other.

and eventually production might move to other (overseas) locations for instance in an attempt to minimize labor and distribution costs;

2. In phase 2, subsequently production starts in places outside the U.S., e.g. in Europe. These production facilities are able to supply their own domestic demand while the U.S. continues to export but now to LDC's in which no production takes place yet;
3. Phase 3, is the phase in which U.S. exports to LDC's have been displaced by exports from other (European) countries who (still) have a cost-advantage in production relative to the U.S.;
4. In phase 4, ultimately the cost advantages of overseas (European) production facilities lead to domestic U.S. production being displaced by exports from the overseas countries;
5. Finally, in phase 5, the product has reached maturity and LDC's export the by now completely standardized product to the United States.

The link between this geographical life cycle and the standard life cycle discussed above is not obvious because the demand aspect is completely absent in this life cycle of location of production and trade. It is, however, easy to see that the actual production, as opposed to demand, in the United States does follow the pattern of the standard life cycle with a first (development) phase in which the product is introduced and total output (and combined world-wide) demand grows slowly and a final (obsolescence) phase during which total production (in the United States) declines slowly and asymptotically to its final level (e.g. zero). Furthermore, production in other advanced countries will follow the same path, only with a lag vis-à-vis the U.S.

6.1.3 The evolution of industries

The evolution of industries, i.e. the development of industries over time, is without a doubt the most widely studied context in which the life-cycle concept has been placed. In seminal studies Gort and Klepper (1982) and Klepper and Graddy (1990) find that a large number of different industries all follow roughly the same pattern of development.[137] Their main focus has been on market structure and they find that industry variables like entry, exit, and the total number of firms active in the industry, follow the familiar hump-shaped path as depicted in Figure 21.

[137] For more work, theoretical, empirical or a combination of the two, on the evolution of industries in general and in the context of a life cycle especially, see Jovanovic (1982), Jovanovic and Saul (1989), Jovanovic and MacDonald (1994a and 1994b), Klepper (1996), Agarwal and Gort (1996), and Agarwal (1998).

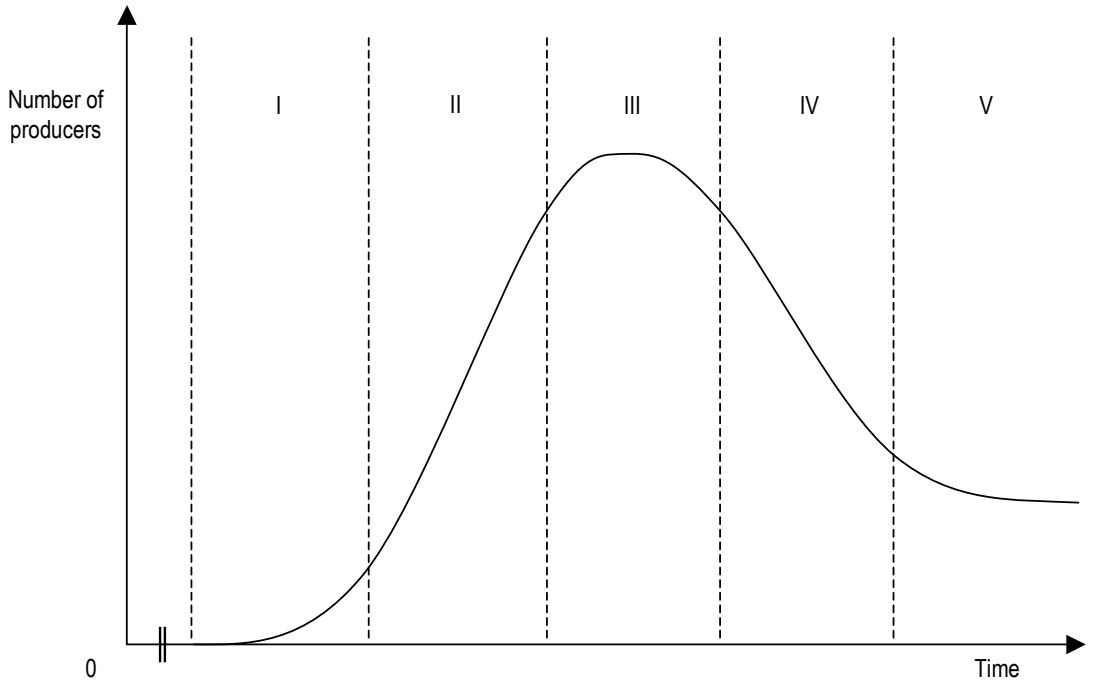


Figure 21, Entry, exit, and the total number of producers over the life cycle.[138]

That is, over the life cycle of an industry, entry number, exit numbers and the total number of firms active in the industry, all follow a pattern of take off, stabilization, and ultimately decline.

Gort and Klepper describe these five stages as follows (p. 631):

I. The first stage starts with the commercial introduction of the product by its very first producer and ends with high rates of entry of new producers;

II. In stage II the sharp increase in the number of producers continues;

III. In stage III some sort of temporary equilibrium is achieved in which entry and exit rates are roughly balanced and the peak number of producers is achieved;

IV. The fourth stage is the stage in which exit numbers overtake entry numbers so that there is net exit or negative net entry;

V. Stage V comes at the end of the stage of negative net entry and again depicts some sort of equilibrium with entry and exit rates roughly balanced.

Besides data on negative net entry, the industry evolution approach to the life cycle has investigated further empirical regularities associated with the life cycle and various stages. The findings, including the one regarding entry and exit are shown in Table 8.

[138] figure adapted from Gort and Klepper (1982, p. 639).

Table 8: Industry evolution.[139]

Stage	Average stage duration	Mean annual net entry	Percentage change in output	Mean annual number of minor innovations	Mean annual number of major innovations	Percentage change in real price	Mean annual rate of patenting
I	14.4	0.50	56.6	0.24	0.24	-13.6	6.6
II	9.7	5.67	35.1	0.25	0.29	-13.0	16.4
III	7.5	0.13	12.3	0.47	0.28	-7.2	26.6
IV	5.4	-4.84	8.1	0.22	0.24	-9.0	21.9
V	…	-0.47	1.0	0.18	0.26	-5.2	7.0

Remarks: data is averaged over 46 new product introductions, which were chosen on the basis of three criteria, (1) to allow sufficient diversity by including consumer, industrial, and military products, (2) to include only products that were 'basic' innovations, and (3) to include products with adequate data on net entry (Gort and Klepper, p. 638). The initial dates of commercial introduction span a 73-year period, beginning with phonograph records in 1887 and ending with lasers in 1960 (ibid.).

Table 8 provides a lot of interesting information especially regarding the development of price and output over time.

The average price of the products examined declines monotonically over time (see the seventh column of Table 8). This in turn means that output measured in absolute units follows a different path than output measured in monetary units. Combining this finding with the idea that diffusion of numbers of products follows the logistic pattern familiar from Paragraph 6.1.1, the result is that demand/output measured in monetary units follows a hump-shaped pattern for demand/output. Assuming that the price of a product declines in a simple exponential pattern this pattern for output/demand can subsequently simply described by a straightforward adaptation of equation (6.1):

$$\text{Output}(t) = P(t) \times \text{price}(t) = \frac{K}{1 + e^{-(a+bt)}} \times P_0 e^{-ct} \tag{6.4}$$

where P_0 is the price of the product at the time of its introduction and c determines the speed at which its price declines. Hence, a link between the industry evolution approach life cycle and the standard life cycle is apparent (see also Appendix I).

The most interesting data depicted in Table 8, however, is the data on growth of industry output (see the third column of Table 8). Gort and Klepper find that rate of output growth declines steadily over time and through the various stages, from an impressive 56.6% to a meager 1.0%. In much the same way as above, this finding also provides a link from the industry evolution approach to the life cycle to the standard approach to the life cycle. When

[139] Source: Gort and Klepper (1982).

the growth rate of output of a product is larger than the growth rate of the whole economy, output of this product as a share of total output increases. When, however, the growth rate of output of a product is smaller than the growth rate of the whole economy, output of this product as a share of total output decreases. This in turn constitutes the familiar hump-shaped life-cycle pattern where the turning-point is reached when the growth of output of the product is equal to the growth rate of total output in the economy (see also Appendix I).
A final interesting observation not reported in Table 8 is that the average length of the introduction stage (I) has seen a remarkable decline over time: for products introduced before 1930 this stage spanned 23.1 years, for products introduced between 1930-1939 this stage spanned 9.6 years, and for products introduced after 1940, it spanned 4.9 years (Gort and Klepper, 640). This strongly suggests that knowledge regarding new products diffuses faster reducing the time to imitation (ibid.).

6.2 Demand and growth without scale effects

6.2.1 Hierarchic demand, saturation, and nonhomothetic preferences

Hierarchic demand, saturation, and nonhomothetic preferences are closely related. The idea is that a person can only consume so much of a certain product before some sort of saturation sets in and the contribution of consumption of additional units of the product to well-being starts diminishing rapidly. The logic of this idea is very present in everyday life, where only few households have more than one lawn mower, or more than one doorbell, simply because the marginal contribution of a second (extra) lawnmower or doorbell is limited. From an economic point of view, however, saturation is most present in the famous law of Engel which states that "the share of food in total expenditures is inversely related to the household's income (or some other measure of its total resources)." (Houthakker 1987b, p. 143; see also Chapter 4), the logic of which was also known by Adam Smith (1979 | 1776) where he notes that "The desire for food is limited in every man by the narrow capacity of the human stomach […]" (p. 269) (see also Falkinger and Zweimüller 1996). Because of saturation, consumer preferences are nonhomothetic, which means that demand does not increase with the same amount for all products when income increases but rather the composition of demand as measured by the shares of income spend on different (types of) products changes with changes in income. In addition, the diminishing contribution to overall

well-being of additional units of a certain product gives rise to a hierarchy of demand (cf. Maslow's 1970 'hierarchy of needs'). The idea of this hierarchy is that some human desires are more pressing than others: the very basic needs such as water, food, sleep, etc. (physiological needs) are the most pressing needs, while self-development and the thriving to exploit ones potential to the fullest (self-actualization needs) are the least pressing needs (Maslow 1970). In the hierarchy, the physiological needs have the highest priority before needs like self-actualization and hence people only consume products that satisfy lower priority needs if the more basic needs have already been satisfied. That is, demand is subjected to hierarchy in which demand shifts from higher priority goods to lower priority goods as soon as additional consumption of products higher in the hierarchy contributes less to well-being than does consumption of products lower in the hierarchy.

The idea that demand develops according to a hierarchy (i.e. saturation and nonhomothetic preferences) has important implication for the study of R&D-based growth (with or without scale effects). The two most important ones are that:

1. It calls for a modelling approach that goes beyond the Dixit-Stiglitz model of monopolistic competition because in their model preferences are homothetic while, as already mentioned more than once in this thesis, a lot of empirical evidence points to the nonhomotheticity of preferences (cf. Engel's law, Deaton and Muellbauer 1983 and Houthakker 1987b);
2. Technological progress, i.e. the introduction of improved and especially new products, occurs along the hierarchy of demand. That is, new goods are either better at satisfying a certain need than the already existing goods, or are able to satisfy a need that until recently was not being satisfied. This in turn relates to the familiar idea of a product life cycle familiar from the discussion of saturation and nonhomothetic preferences in Chapter 4 and of course the discussion in the previous section. In terms of hierarchic demand the life cycle is best understood as a good that once was a luxury, over time, as income rises, evolves into a necessity. Examples of this are plenty, e.g. refrigerators, mobile phones, magnetrons, etc. What is more, this evolution from luxury to necessity can be seen as an additional manifestation of creative destruction, besides the three discussed in the previous chapter. The forces underlying this form of creative destruction, are probably habituation, the relativity of aspiration and related phenomena (e.g. Hirsch 1976, Scitovsky 1976, Abramovitz 1979, and Easterlin 2001): over time people get used to luxuries and are no longer able to live their life without them. In addition, a black-and-white TV does not seem to be such a luxury when LCD screens are available as well.

6.2.2 Theoretical work with hierarchic demand

The idea of demand developing along a hierarchy is used in various theoretical work. Some of this work has already come up though only briefly in previous chapters, e.g. Echevarria (1997), and Falkinger and Zweimüller (1996 and 1997). Other oft-mentioned work is, amongst others, Murphy, Shleifer and Vishny (1989), Zweimüller (2000), Föllmi and Zweimüller (2000 and 2002), and Matsuyama (2000 and 2002).

Interesting is that this work uses nonhomothetic preferences to study different though related phenomena.

For instance, Murphy, Shleifer and Vishny (1989) use hierarchic demand to study the effect of inequality in the income distribution on growth through industrialization, The idea is that a country can grow and deindustrialize successfully only if the domestic markets are large enough. Only if these markets are large enough, will it be profitable for firms to engage in R&D. An unequal income distribution may mean that only a relatively small upper-class can afford new products, rather than a relatively large middle-class, effectively inhibiting R&D as no profit can be made. Matsuyama (2000) on the other hand uses the nonhomothetic specification of preferences in a Ricardian model of trade. The idea in this study is that homothetic preferences (e.g. Dixit and Stiglitz) are not appropriate for the study of issues like the geographical life cycle (e.g. Vernon 1966, Hirsch 1967 and Wells 1972a) and the deterioration of the terms of trade for poor countries, i.e. the Prebisch-Singer hypothesis[140]. Most interesting, however, at least for the pervading issue of growth scale effects, are the studies that have used hierarchic demand in a model of growth, e.g. Föllmi and Zweimüller (2000 and 2002), Zweimüller (2000), and Matsuyama (2002).

Of these in turn, Matsuyama (2002) is most interesting for the clear relation it presents between hierarchic demand on the hand and the product life cycle on the other. In his model demand for products takes off continuously to form the Flying Geese pattern[141] depicted in Figure 22 below.

---

[140] See Prebisch, R. (1950), *The Economic Development of Latin America and Its Principal Problems*, New York, United Nations, and Singer, H.W. (1950), 'The distribution of gains between borrowing and investing countries', *American Economic Review, Papers and Proceedings*, 40 (2), 473-85.

[141] Matsuyama (2002) borrows the idea to describe the pattern of life cycles taking off one after another as the Flying Geese pattern, from Akamatsu (Akamatsu, K. (1961), 'A Theory of Unbalanced Growth in the World Economy', *Weltwirtschaftliches Archiv*, 86 (2), 196-215.) who uses it to describe the time pattern of imports, domestic production, and exports of manufacturing goods (Matsuyama 2002, p. 1037; see also Paragraph 6.1.2 on the geographical life cycle).

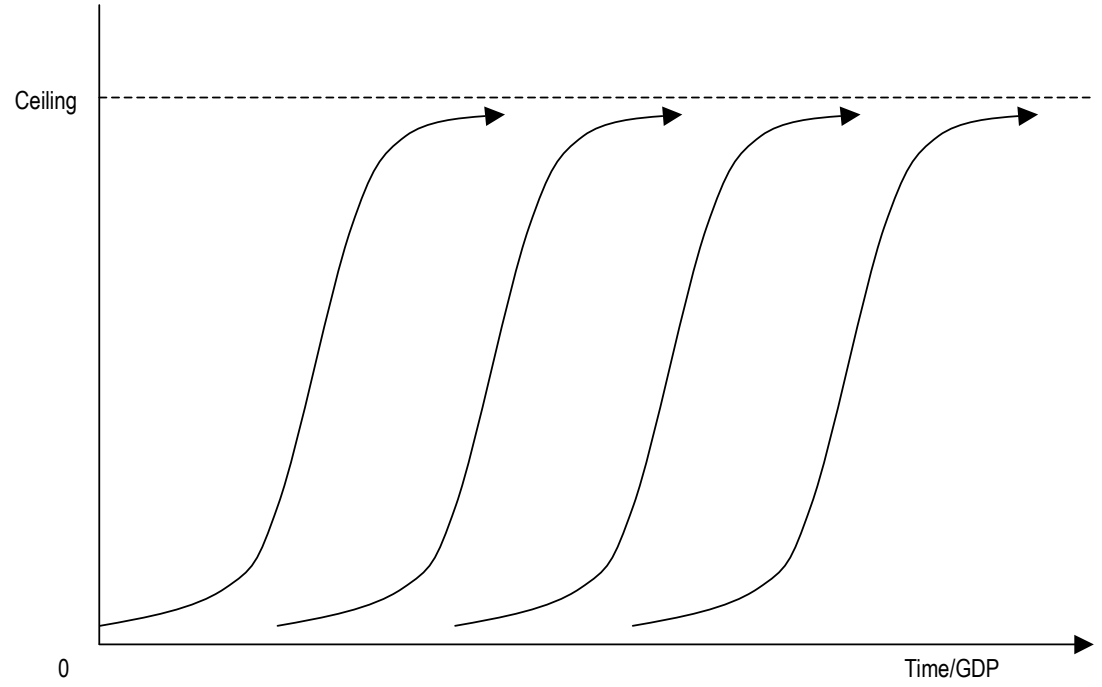


Figure 22, The Flying Geese pattern of different product life cycles.[142]

That is, in Matsuyama's (2002) model new products are invented and introduced to the market continuously over time. The Flying Geese pattern that results is sustained by a two-way causality between productivity improvement and consumption shifting to new goods. Hence, the model is very similar to - though immensely more elegant than - that of Chapter 4, in which the intricate relation between consumer preferences, income, relative prices, technological progress etc., was assumed to give rise to some sort of feedback-loop between income and growth. In Matsuyama's (2002) model, there is a similar feedback-loop in the form of a trickle-down and a trickle-up process in which hierarchic demand and nonhomothetic preferences are important. Starting point is that consumer's differ only with respect to their level of wealth (income) and the nonhomotheticity of preferences subsequently implies that demand for a new good, i.e. a good that is lower in the hierarchy than the good presently consumed, takes off very slowly at first because it is only consumed by the happy few. When the richer consumers buy a new low priority product they start the trickle-down process however. This is a process of product-specific learning by doing which leads to productivity improvements, which in turn lead to a lower price of the product and in this way enables poorer households to buy the good as well. The trickle-down process, is thus simply consumption of a (new) (low priority) good by (high-income) households enabling

[142] See Matsuyama (2002).

other (low-income) households to consume the product as well. The outcome is that production of - and demand for - the product takes off.[143] The purchase of the good by other (low-income) consumers following the trickle-down process leads to even more product-specific learning by doing, driving down the price of the new good even further. This in turn enables the early consumers of the product to expand their consumption to the next product in the hierarchy, and this is the trickle-up process. The trickle-up process leads to the start of a new trickle-down process and hence the result of the combination of a trickle-down and a trickle-up process is the Flying Geese pattern of the continuous introduction of new products each of which go through the familiar life-cycle pattern (see also footnote 143). The model of Matsuyama (2002) incorporates much of the features of the mathematical stylization discussed in Chapter 4, including the very important idea of limited interindustry/intersectoral spillovers, i.e. spillovers are diluted through technological distance and knowledge is product/sector-specific (see also Chapter 3). What is missing in his story, at least from the point of view of this thesis in which R&D-based growth plays such an important role, is an explicit role for R&D either in the form of productivity improvements or in the form of the creation and introduction of new and/or improved products, i.e. a role for technological progress in a broad sense, however. Instead, Matsuyama (2002) relies on learning by doing for increases in productivity and ultimately income and simply assumes that new goods, which are lower in the hierarchy, are already exogenously available and await actual consumption.

The models developed by Föllmi and Zweimüller (2000 and 2002) have much the same features as that of Matsuyama (2002) with the added bonus of endogenized R&D-based technological progress. In their models of R&D-based growth, endogenous R&D-based productivity growth and endogenous R&D-based product innovation, does come at a cost, however, which is that in their models growth just like in the early models of R&D-based growth (e.g. Romer 1990, Grossman and Helpman 1991 and Aghion and Howitt 1992) does exhibit a scale effect. As a way to dispose of this (unwanted) feature of the model Föllmi and Zweimüller (2002) propose the use of the familiar fishing-out mechanism (cf. Jones 1995b), which in turn has the negative side effect of leading to a semi-endogenous model of R&D-based growth (see also Aoki and Yoshikawa 2002).

The mathematical stylization discussed in Chapter 4 and Matsuyama's (2002) model have, however, shown that dilution of spillovers can achieve much the same effect of absence of

[143] Because at the same time prices decline, Matsuyama's (2002) model accurately describes the beginning of the industry life cycle depicted in the third and seventh column of Table 8.

growth scale effects. The conclusion ending this section consequently is simply that growth without scale effects can easily be modelled when the early models of R&D-based growth are equipped with hierarchic demand (nonhomothetic preferences and saturation) and dilution of spillovers (through technological distance) in much the same way as is dome in the mathematical stylization. In other words, merging a model like that of Peretto and Smulders (2002), dilution through technological distance, with that of Matsuyama (2002), hierarchic demand, renders an elegant framework for modelling growth without scale effects.

6.3 Discussion and conclusions

In Chapter 2 and Chapter 3, two important and closely related conclusions regarding the modelling of growth (without scale effects) were made:

1. a model of growth that is not only consistent with the absence of a scale effect in growth but also with other empirical facts is preferred over a model that lacks such a foundation in observed economic reality,
2. incorporation of observed economic facts and phenomena can actually provide important building blocks in a model of growth without scale effects.

The conclusion of the previous section can best be seen as a combination of these two earlier conclusions: it presents two aspects of economic reality that hold the key to the understanding of - and the ability to credibly model - growth without scale effects, i.e. (empirically observed) dilution (through technological distance) and (empirically observed) hierarchic demand.

This of course is a very important conclusion and indeed one of the most important conclusions of this thesis. The most important conclusion, however, goes one step further to state that not only growth without scale effects but also - and very importantly so - other real-life phenomena like structural change and product (/industry/etc.) life cycles can all somehow be traced back to hierarchic demand and dilution through technological distance, in turn also real-life phenomena.

Throughout this thesis the mechanism by which the economy ‘achieves’ the absence of growth scale effects (where logically such effects are expected as the result of knowledge’s nonrivalrous nature) has been described as an intricate relation between factors like consumer preferences, income, relative prices, technological progress and so forth, which gives rise to

various forms of creative destruction. One of the important aspects of this oft-mentioned intricate relation is that products go through a life cycle in which demand first takes off, then stabilizes and ultimately declines to its final level, i.e. there is a hump-shaped relation between demand for a product measured as a share of total production on the one hand and time or total production on the other. This leads to the idea that the intricate relation and the various manifestations of creative destruction associated with it can better - and more insightfully - be described as an 'endogenous product life cycle'. This term best captures the idea that due to a combination of circumstances, with endogenous R&D-based technological progress in a broad sense being the biggest culprit, over time older products are replaced by new products better able to satisfy the needs of consumers. In addition, it clearly states the way in which growth without scale effects (and other important phenomena like structural change) can be credibly modelled, namely by endogenizing the life cycle for instance by using nonhomothetic preferences (see also Section 6.2). Finally, the idea of an endogenous life cycle relates back to the idea of endogenous network participation discussed in Chapter 3. Indeed, it seems that hierarchic demand (saturation and nonhomothetic preferences) can be incorporated in the Peretto-Smulders model to create markets which differ in the opportunities they offer to firms, especially in the medium to long run. Combining an endogenous life cycle with endogenous network participation, could have the added benefit of not only explaining macro level phenomena like growth without scale effects and structural change, but also provide a framework to analyze micro-level phenomena like diversification and its performance effects (see also Matsusaka).

7 Concluding remarks

The aim of thesis has simply been to gather clues that would help further the understanding of (the mechanism underlying) the observed absence of growth scale effects, i.e. the observation that larger economies (suitably measured) do not grow faster than smaller economies (suitably measured) do. The idea was that existing models of R&D-based models of growth without scale effects suffer from the Solow critique (Solow 1994, p. 53): they make favorable assumptions regarding the extent of the spillovers associated with knowledge's nonrivalrousness, which are the driving force behind the predict of growth scale effects of the early models of R&D-based growth (e.g. Romer 1990, Grossman and Helpman 1991, and Aghion and Howitt 1992), but often the much-needed (micro-)economic foundation for these crucial assumptions is lacking (cf. Solow 1994, p. 53). Assuming that knowledge is rivalrous (as opposed to nonrivalrous) to limit spillovers so as to cause the absence of growth scale effects does not shed much light on the issue of growth without scale effects however. The broader idea of this thesis consequently was that these clues could help enhance the ability to credibly model growth without scale effects effectively rebutting the Solow critique.
Clues have been gathered by taking a closer look at various empirical regularities at both the micro and the macro level, expected to be related to factors important in growth without scale effects in general and knowledge, R&D and spillovers in particular. In doing so indeed several important aspects of economic reality surfaced that could and should play an important role in explaining the absence of a scale effect in growth where logically such an effect would be expected as the result of knowledge's nonrivalrousness.

More concrete the clues all point to a single, though broad, mechanism underlying the observed absence of a scale effect in growth (and also other observed economic phenomena), namely an intricate relation between factors like consumer preferences, income, relative prices, technological progress etc., which gives rise to various forms of creative destruction (and ultimately leads to growth without scale effects).
Because one of the most important effects of the intricacies in the relations between factors like consumer preferences, income, relative prices, technological progress and so forth, is that products go through a life cycle in which demand first takes off, then stabilizes and ultimately declines to its final level, part of this mechanism can best be described as an 'endogenous product life cycle'.

This is so because the term endogenous life cycle best captures the idea that factors like income, relative prices, technological progress affect each other 'endogenously', i.e. in complex (and intricate) ways in which two-way causality plays an important role. For instance, R&D efforts and consequent technological progress in the form of an increase in the level of productivity of a certain sector not only effects the level of income, but also relative prices and demand, which in turn provide incentives for future R&D activities and consequent technological progress. That is, product life cycles are created in the most endogenous way one could imagine, and exogenous factors play the smallest possible role in the fact that demand for a product goes through stages of take off, stabilization, and ultimately decline. Although obviously several factors play a role in 'the endogeneity of the product life cycle' as here described, it seems possible to trace the root of this endogeneity back to a two-way causality between demand (in abroad sense) and technological progress in a broad sense. Demand factors, or more broadly individuals' well-being functions, provide incentives and influence firms' R&D decisions, e.g. the amount of R&D to perform and the direction of it. Technological progress, i.e. productivity growth and the introduction of new and/or improved products, is thus affected by demand, but in turn affects demand and, indeed, future R&D activities, itself.

The endogenous life cycle described here constitutes only part of the intricate relation between factors like consumer preferences, income, relative prices, technological progress and so forth, mentioned above, and consequently constitutes only part of the mechanism stated above to give rise to various forms of creative destruction and ultimately lead to the absence of a scale effect in growth.
The explicit independent attention it has received here is warranted by the fact that models exist that actually model the endogeneity of life cycle, i.e. the two-way causality between demand on the one hand and technological progress on the other. In Matsuyama (2002)

technological progress, i.e. productivity growth, resulting from product-specific[144] learning by doing and hierarchic demand (saturation and nonhomothetic preferences) combine to make a model that (i) describes the Flying Geese pattern of the continuous introduction of new products each of which go through the familiar life-cycle pattern, and (ii) does not have the prediction of a scale effect in growth. More interesting, given the subject of this thesis, subsequently are the models developed by Föllmi and Zweimüller (2000 and 2002). These models have much the same features as the Matsuyama-model with the attractive additional feature of endogenized R&D-based technological progress. Unfortunately, their models just like the early models of R&D-based growth (e.g. Romer 1990, Grossman and Helpman 1991 and Aghion and Howitt 1992) predict that growth exhibits a scale effect. As a way of disposing of this unwanted feature of their models, Föllmi and Zweimüller (2002) subsequently propose the use of the familiar fishing-out mechanism (cf. Jones 1995b). This seems to be a big disappointment, because (1) this mechanism seems to be, to paraphrase Solow (1994), the kind of (unearned) favorable assumption this thesis would like to render unnecessary in the modelling of R&D-based growth without scale effects, and because (2) the result is semi-endogenous, as opposed to fully, endogenous growth, which is not in line with the spirit of endogenous (R&D-based) growth, which is to address the unwanted prediction of exogenous long-run growth of the neoclassical model of growth (e.g. Solow 1956).
However, the theoretical work of Matsuyama (2002) and Föllmi and Zweimüller (2000 and 2002), also allows the drawing of two conclusions, which are far from disappointing:

1. dilution of knowledge through technological distance, i.e. knowledge is (at least partly) product-specific, is important for the absence of a scale effect in growth,
2. the intricacies in the relation between factors like consumer preferences, income, relative prices, technological progress and so forth, i.e. the interaction between technological

[144] Learning by doing is by its very nature closely tied to a single product or class of products. Nonetheless, the assumption Matsuyama (2002) implicitly makes is that knowledge is diluted through technological distance, i.e. knowledge is less useful when used in the production of goods other than the one(s) for which it was originally conceived (cf. Peretto and Smulders 2002).
The effect of dilution of knowledge through technological distance is that (product-specific) knowledge is likely to become obsolete over time when the goods in the production of which it was used are no longer in demand. For so far as the decline in demand is the result of endogenous technological progress, the rate of knowledge obsolescence is also endogenous (see Chapter 5).
The idea of knowledge dilution originated in the work of Adams and Jaffe (1996) who found that the contribution of knowledge (R&D) to plant-level productivity was diminished when spread over a larger number of plants. They subsequently suggested technological distance, a concept introduced by Griliches (1979) and operationalized by Jaffe (1986), to provide a possible explanation for their findings. Finally, Peretto and Smulders (2002) have implemented the idea of knowledge dilution through technological distance in a model of growth without scale effects, not without making some assumptions that cannot withstand the Solow critique (Solow 1994) regarding unearned favorable assumptions however.

progress on the one hand and individuals' well-being functions (demand) leading to creative destruction manifesting itself only in the existence of product life cycles, i.e. in the replacement of one good by another and in demand for a product going through stages of take off, stabilization and ultimately decline, are not 'powerful' enough to cause the absence of growth scale effects.

Hence, the premise that started this passage, namely that the endogenous life cycle, i.e. the two-way causality between demand (or better individuals' well being functions) on the one hand and technological progress on the other, constitutes only part of the mechanism giving rise to various forms of creative destruction and ultimately lead to the absence of a scale effect in growth, is confirmed.

It would still require a strong leap of faith, however, to conclude almost solely on the basis of the studies by Matsuyama (2002) and Föllmi and Zweimüller (2000 and 2002), that combined dilution through technological distance and the intricate relation between consumer preferences, income, relative prices, technological progress etc., lead to growth not exhibiting a scale effect. Nonetheless, this is a justifiable conclusion, for it is based not only on this theoretical work, but also on several other researchers' findings, ideas and issues, discussed, forwarded and raised in this thesis, the most important of which are repeated here:

- Chapter 4 has developed a mathematical stylization in which demand shifting from 'older' to 'younger' sectors, in turn the result from the intricate relation between consumer preferences, income, relative prices, technological progress etc., combines with productivity differences between the sectors ($TFP_{'old'} > TFP_{'young'}$), in turn the result of dilution of spillovers associated with knowledge's nonrivalrousness through technological distance, to cause the absence of scale effects at the aggregate level (while such effects may be existent at the sectoral level).
- Similarly, Chapter 5 has discussed that there are more intricacies in the relation between consumer preferences, income, relative prices, technological progress and so forth, than just the ones manifesting themselves in the endogeneity of the product life cycle, or stated differently, creative destruction manifests itself not only in the form of demand for a product following a pattern of take off, stabilization, and ultimately decline. Take as astartin point two interesting findings regarding the relation between income an well-being (which consumers are expected to maximize through consumption of goods in a way determined by their individual well-being function) have been reported in literature: (1) for well-being absolute income matters less than relative income (see for example

Easterlin 1974 and McBride 2001), and (2) in periods with increasing income, an increase in well-being was lacking (see for example Easterlin 1974 and Blanchflower and Oswald 2003). On the basis of these puzzling findings researchers have concluded that there is a negative externality of consumption (Layard 1980) and that the case for growth as a way of increasing well-being is no longer so clear (e.g. Easterlin 1974, Hirsch 1976, Scitovsky 1976, and Easterlin 1995). In Chapter 5, however, the implication of these findings that there is a discrepancy between output (and growth) measured by the extreme measure of (pure) physical output on the one hand, and by the extreme measure of well-being, has been emphasized. In this picture, output measured using the usual monetary unit of account is a combination of these two forms. A factor like (physical) productivity growth, which is on the technological side, meets the complexities of human nature (individuals' well-being functions) in the middle and together they combine to the familiar monetary measures of relative prices, income, TFP and so forth (cf. Pigou 1952, p. 50). This in turn led to the idea that several of the intricacies present in individuals' well-being functions, e.g. saturation, might not reveal itself in physical output measures, but only in monetary output measures. For instance, peoples' tastes (ibid.) are revealed in the (monetary) value people place on certain products, i.e. in the price of these products, and consequently affect TFP measured using monetary units of account. Similarly, (physical) productivity growth reveals itself in the relative price of a product and consequently in TFP. The intricacies of human nature (i.e. of individuals' well-being functions, of taste, etc.) subsequently are such that factors like the social context in which consumption takes place plays a role in the well-being people derive from the consumption of a certain product and hence affect the market price of these products and consequently monetary measures of output and growth. Furthermore, the complexities present in individuals' well-being functions gives rise to additional manifestations of creative destruction. The introduction of new products may for instance diminish part of the ability to provide well-being of the already existing goods, with the logical effect of the value placed on these already existing goods changing downward, in turn leading to a price decrease and quite possibly to changes in measured levels of income and productivity. That is, ultimately manifestations of creative destruction resulting from the intricacies of human nature (i.e. taste, individuals' well-being functions, and so forth) may reveal themselves in normal measures of output and growth. Indeed, productivity growth for a certain product may exhibit a scale effect when measured in pure physical units, whereas the monetary variant, does not due to these intricacies. That such a depiction of reality is not unlikely follows in

turn from the above-presented finding of a discrepancy between observed growth of income and the lack of growth in well-being. Unfortunately, the pure data needed to test this premise is missing.

With the findings, ideas and issues, discussed, forwarded and raised in Chapter 4 and 5, and repeated here in mind, the conclusion that, combined, dilution through technological distance and the intricate relation between consumer preferences, income, relative prices, technological progress etc., cause the absence of growth scale effects, seems more than justifiable. It subsequently is the most important conclusion of this section and probably also this thesis.

Nonetheless, some other issues arise, especially regarding directions for future work.

1. The several intricacies present in peoples' well-being functions discussed above and more at length in Chapter 5, need not be taken into account explicitly in a model of R&D-based growth without scale effects. Indeed, a framework combining dilution through technological distance (cf. Peretto and Smulders 2002) with hierarchic demand (nonhomothetic preferences and saturation; cf. Matsuyama 2002) will easily sustain a credible model of growth without scale effects firmly rooted in empirical regularities, and able to withstand critique along the lines of Solow (1994). A first direction for future work subsequently is the formalization of such a framework.
2. Other possibly fruitful directions for future work, are many but have already been discussed at some length in conclusions and concluding remarks following various sections and chapters. A second direction that deserves mentioning here nonetheless is to investigate the relation between firm size and productivity growth to check for firm-level growth scale effects (where the combination of dilution through technological distance and the intricate relation between consumer preferences, income, relative prices, technological progress and so forth, is expected to cause such effects to be absent perhaps even more so than at the aggregate level).

Appendix A Rent spillovers and pure knowledge spillovers

This appendix briefly extends the discussion of the two types of spillovers related to knowledge and R&D, and the use of weighting schemes to distinguish between them, already presented in Chapter 2 (see also Griliches 1992 which presents a discussion of rent spillovers and pure knowledge spillovers similar to the discussion presented here).

As stated, there are two types of spillovers related to knowledge and R&D, rent spillovers and pure knowledge spillovers (cf. Scitovsky 1954).

1. Rent spillovers, result from the fact that a firm is unable to appropriate fully the rents from its innovation(s), new (firm-specific) knowledge taking the form of (embodied in) a new or improved product. This is due to the fact that the firm is unable to perfectly price discriminate and to the fact that there are other firms supplying (close) substitutes. Thus, new knowledge increases surplus only part of which ends up in the hands of the innovating firm in the form of producer surplus while the rest ends up in the hands of buyers in the form of consumer surplus. When such an innovation created in industry i is subsequently used as an input in downstream industry j, it may contribute to productivity growth in this last industry (downstream). That is, part of the overall benefits of an innovation may spillover from one industry to another. However, as Griliches (1992) notes, this type of spillovers are not real (pure) spillovers (or externalities), but rather they are pecuniary externalities (see also Scitovsky 1954). These type of externalities would not exist if the complete quality adjusted price of new and/or improved products could be determined accurately, for then productivity increases could be attributed to technological progress in the correct industry, i.e. to the industry where it actually originated (see also Griliches 1979). What is more, rent spillovers are not the type of externalities that play such a large role in the R&D-based models that are of interest in this thesis (e.g. Romer 1990, Grossman and Helpman 1991, Jones 1995b, Smulders and van de Klundert 1995, Young 1998, Li 2000, and Peretto and Smulders 2002).
2. Pure knowledge spillovers, or non-pecuniary externalities as Griliches (1992) calls them, on the other hand, are directly related to the nonrivalrous nature of knowledge, which in turn is the ultimate driving force behind a possible scale effect in growth. With this type of externalities, knowledge is not embodied in products or services as is the case with rent spillovers, and consequently no market transactions are required for these externalities to

occur. Rather, pure knowledge spillovers result from disembodied knowledge which nonrivalrousness leads to increasing returns and from knowledge flowing (almost) freely between firms, sectors, regions, etc. That is, a number of firms (or industries, etc.) get hold of a single 'piece' of knowledge, for instance through coverage in written press or through interactions with other firms, which they all can use productively, i.e. which increases their productivity. Ultimately knowledge's nonrivalrous nature and the pure knowledge spillovers associated with it might lead to a scale effect in growth. Hence, in this thesis in general and in this chapter and section in particular the interest goes out to pure knowledge spillovers.

For the purpose of this thesis, distinguishing between the two types of spillovers is thus important. Unfortunately, this is also difficult. However, in empirical work a solution to this problem has been created in the form of weighting schemes. Weighting schemes are used because, the effect of industry i's stock of accumulated R&D investment on productivity growth in industry j is moderated by 'how much' of this knowledge stock actually spills over from industry i to industry j (or from region i to region j, country i to country j, etc.). In the literature on spillovers, two types of weighting schemes have been widely used.

1. Weighting the contribution of other industry's knowledge stocks to productivity in a certain industry by correcting them for input-output relations between the several industries, e.g. trade shares (see for instance Scherer 1982). Undoubtedly inspired by literature that stressed the role of trade in the growth process (e.g. Grossman and Helpman 1991), a lot of studies (e.g. Coe and Helpman and Coe, Helpman and Hoffmaister), have used this type of weighting scheme to assess the impact of international spillovers. However, as noted by Griliches (1979), using a weighting scheme based on input-output relations, has the effect of creating a framework in which the spillovers investigated are rent spillovers. This is because using input-output relations makes the assumption that spillovers are somehow associated with the flow of goods and/or services between industries and in the framework thus created, the contribution to productivity growth is attributed to industries in a manner that has much similarities to the way in which prices can (and should) be corrected for quality increases.
2. The second type of weighting scheme tries to correct for the actual usefulness of industry i's R&D (capital) for industry j by taking into account the technological distance between the two industries (see Griliches 1979). That is, the usefulness of knowledge spillovers between two industries that are technologically related will be higher than between

industries that are technologically far apart. Hence, using a measure of technological distance to correct for the actual usefulness of knowledge emanating from industry i for industry j, naturally creates a framework in which the spillovers studied are (pure) knowledge spillovers. In the words of Griliches (1979), "True spillovers are the ideas borrowed by the research teams of industry i from the research results of industry j. It is not clear that this type of borrowing is particularly related to input purchase flows. The photographic equipment industry and the scientific instruments industry may not buy much from each other but may be, in a sense, working on similar things and hence benefiting much from each other's research." (p. 104). Verspagen (1997) and Frantzen (2002) are some of the studies that have used a technological distance weighting scheme in their investigation of both international and intersectoral spillovers. The study of Jaffe (1986) is one of the firsts to operationalize Griliches' (1979) concept of technological distance, and apply it to micro/firm-level data on amongst others R&D and productivity.

Appendix B: Firm size and growth

This Appendix Discusses in some detail several studies that have examined the relation between firm size and firm growth, namely Mansfield (1962), Kumar (1985), Evans (1987a and 1987b), Hall (1987), Dunne, Roberts and Samuelson (1989), and Dunne and Hughes (1994).

Mansfield (1962)

Mansfield investigates whether Gibrat's law of proportionate effect describes the process of firm growth well. Gibrat's law states that the growth of firms is independent of their size and that thus a firm with ten million dollars in sales or 100 employees, is equally likely to double in size over a given period than is a firm with one billion dollar in sales or 10,000 employees. Using a small data set with data for the steel, petroleum, and rubber tire industries, he finds that Gibrat's law is not an accurate description of growth of firms. This finding is robust to various adjustments to the data set, e.g. leaving out firms that have left the industry and limiting it to contain only firms beyond a certain threshold size. Mansfield further investigates the effect successful innovation has on firms' growth rates and finds that, on average, successful innovators grow twice as fast as other firms ceteris paribus. The effect of successful innovation differs, however, with firm size and for smaller firms the growth difference between successful innovators and other comparable firms is larger than it is for larger firms.

Kumar (1985)

Kumar investigates the relationship between size, growth, and acquisition activity in a sample of quoted companies in the United Kingdom for the period 1960-1976. This study uses a data set which contains five measures of size, net assets (total assets less current liabilities), physical assets (tangible fixed assets plus stocks plus work in progress), equity assets (total issued capital plus reserves), the number of employees and total sales, and two measures of growth, total growth (size at time t + 1divided by size at time t) and growth by acquisition (expenditure on acquisitions divided by opening size as measured by net assets). Kumar finds that size and firm growth are weakly negatively related, irrespective of, the size measures used (net assets, physical assets, equity assets, employees, sales), the period examined (1960-1971, 1966-1976) and/or the industry examined (manufacturing, non-manufacturing).

Furthermore, the study finds a weak negative relation between size and growth by acquisition for manufacturing firms, but for all firms no systematic relation between size and growth by acquisition was found.

Evans (1987a)

Evans (1987a) uses a sample of all firms in 100 manufacturing industries for the period 1976-1980 to examine aspects of industry dynamics, e.g. firm growth and the variability of firm growth, and characteristics of firms, e.g. size and age. The sample is taken from the Small Business Data Base (SBDB) and contains data on firm employment (size), firm age, the number of plants a firm operates and the 4-digit SIC code of a firm's main activity. His study finds several interesting relationships. A first is that (i) there is a negative relation between firm growth and firm size for 89% of the industries, and a negative relation between firm growth and firm age for 76% of the industries. A second interesting relationship is that (ii) the variability in growth rates decreases with age in 85% of the industries and decreases with size for 80% of the industries. The sign of this last relation, however, seems to vary with age and the number of plants a firm operates. Nonetheless, the effects of age are broadly consistent with the model of selection and industry evolution developed by Jovanovic (1982).

Evans (1987b)

Evans (1987b) studies the same relations as Evans (1987a), namely the relation between firm growth and firm size and the relation between firm growth and firm age, for a sample of some 20,000 manufacturing firms for the period 1976-1982, The data is once again drawn from the SBDB so size is measured by the number of employees. He confirms the findings of his earlier work (Evans 1987a) that firm growth decreases with firm age and firm size which is consistent with the model of Jovanovic (1982). In addition, Evans (1987b) examines the relationship between growth and size further and finds that the relationship holds throughout several alternate samples, that the relationship is highly non-linear, and that the relationship is robust to for instance problems related to heteroscedasticity.

Hall (1987)

Hall tests several econometric explanations for the finding of a negative relation between firm size (as measured by firm employment) and growth. She uses a sample of publicly traded firms active in the U.S. manufacturing firms and finds that this relation is robust and not the result of for example measurement error or sample attrition. The smaller firms in Hall's

sample grow faster than larger firms and the difference in annual growth rates between firms in the 25th and 75th size percentile equals 4 percentage points. The growth rates of smaller firms also have higher variance (at least twice as large), than those of larger firms, as was noted by earlier studies (e.g. Evans 1987a). Furthermore, Hall reports on some descriptive regressions pertaining to the relation between (physical) investment and subsequent firm growth, and R&D effort and subsequent firm growth. She finds that an increase in investment of 4 million dollar in 1976 is associated with a 1 percent increase in the annual growth rate for the period 1976-1979 and that the same effect is achieved with a 2 million dollar increase in R&D expenditures.

Dunne, Roberts and Samuelson (1989

Dunne, Roberts and Samuelson examine the pattern of post-entry employment growth of plants in the U.S. manufacturing sector (four-digit SIC code) over the period 1967-1977. They find that large plants have lower growth rates relative to small firms, and that multi-unit plants grow faster than single-unit plants of similar size. In fact, large single-unit plants have negative average growth rates. (In an accompanying paper using much the same data (Dunne, Roberts and Samuelson 1988), they find that on average, across industries, the size of a firm relative to the size of all the firms in the same industry increases with age).

Dunne and Hughes (1994)

Dunne and Hughes repeat much of the work of for instance Evans and Hall using a database of United Kingdom companies rather than U.S. (manufacturing) companies. They find that in their sample of quoted and unquoted firms active in a wide range of industries for the period 1975-1985, smaller firms grow faster than larger firms (as measured by net assets) and that growth and age are negatively related. In addition, they find that these results are not induced by sample problems like selection bias. Hence, Dunne and Hughes confirm the findings of other studies (e.g. Evans and Hall).

Appendix C: Pure knowledge spillovers at the firm level

The first section of this Appendix discusses in considerable detail several studies that have examined interfirm pure knowledge spillovers, namely Jaffe (1986), Jaffe (1988), Megna and Mueller (1991), and Acs, Audretsch and Feldman (1994). In the second section several studies that have examined intrafirm pure knowledge spillovers, namely Adams and Jaffe (1996), Henderson and Cockburn (1996) and Klette (1996), are discussed in the same manner.

C.1 Intrafirm pure knowledge spillovers

Jaffe (1986)

Jaffe investigates two 'supply side' factors that are expected to determine 'the rate and direction of inventive activity' (Nelson 1962), namely technological opportunity and R&D (knowledge) spillovers. The idea is that the productivity of a firm's R&D efforts is affected by technological opportunity, "exogenous variations in the cost and difficulty of innovation in different technological areas" (p. 984), and through the R&D efforts of other companies. The effect of R&D efforts by other firms in turn are affected by the distance in 'technology space', i.e. the closer two firms are in technology space, the more they will benefit from R&D performed by the other.

1. In the first step of his research, Jaffe operationalizes Griliches' (1979) concept of technological distance (or equivalently technological proximity). For this, he starts with determining a firm's technological position. A firm's technological position is described by how the firm's R&D expenditures are spread over K areas of technology. The technological position of a firm can then be denoted by a vector $F = (F_1, \ldots, F_K)$ where F_k is the fraction of the firm's R&D budget devoted to area k (p. 985). Unfortunately, there is no data available on how firms distribute their R&D budget over different technological areas. There is, however, data available on the different areas of technology to which (some of) the results of firms' R&D efforts, patents, belong. The patent office assigns each patent granted to a firm to a certain technology-based patent class (p. 985). Jaffe subsequently uses this data on patents to determine firms' technological position using a vector containing 49 elements (49 patent classes). The technological proximity (distance) between two firms i and j is subsequently calculated as the angular separation of the vectors F_i and F_j. This proximity measure in turn is used in the creation of a 'spillover

pool', which is a summation of all other firms' R&D expenditures weighted by technological proximity. In addition, the technological position vector is used to assign a firm to a certain technological cluster with the idea that firms in the same cluster face similar technological opportunities. In total, firms where assigned to 21 clusters, which remained from an original selection of 25 clusters based on the 25 largest patent classes).

2. Jaffe's second step is to develop estimating equations for firms' (i) R&D output, (ii) profits, and (iii) market value. He models R&D output as a function of its R&D efforts, the product of the total potential spillover pool and its R&D efforts (which of course reflects the idea that a firm is better able to benefit from the available spillover pool the more R&D it performs itself), the firm's total potential spillover pool, and its technological opportunity as indicated by a dummy variable indicating the firm's technological cluster. Because R&D output is ultimately measured by the number of patents a firm generates, the estimating equation is further corrected for the firm's propensity to patent. Firm profits are modeled as a function of the firm's market share, its stock of accumulated R&D investments, the product of the firm's spillover pool and its stock of accumulated R&D investments (which of course reflects the idea that a firm is better able to benefit from the available spillover pool the more R&D it has performed in the past), its capital stock, the spillover pool, the four-firm concentration ratio of the firm's industry, and its technological opportunity (where the firm's market share and the four-firm concentration ratio are weighted averages for all the industries in which the firm is active). Finally, market value (measured by the logarithm of Tobin's q (cf. Tobin 1969)) is modeled as a function of the firm's stock of accumulated R&D investments divided by its capital stock (its R&D-capital ratio), the product of the spillover pool and its R&D-capital ratio, its market share, the spillover pool, the four-firm concentration ratio, and its technological opportunity.
3. In the third step, these three equations are estimated using cross-sectional data from the National Bureau of Economic Research (NBER) R&D panel. Regarding the model of R&D output, Jaffe finds statistically significant results for all elasticities and coefficients. The elasticity with respect to own R&D is 0.875 plus 0.352 times the logarithm of the firm's spillover pool. Since the logarithm of the pool is constructed to have zero mean, the elasticity is 0.875 for the average firm, and higher (lower) for firms with above (below) average spillover pools (p. 994). Similarly, the elasticity of the spillover pool is 0.509 plus 0.352 times the logarithm of the firm's R&D efforts. For the firm with mean logarithm of R&D effort, the elasticity of patents with respect to the R&D performed by other firms is

approximately 1.1. If all firms increased their R&D efforts by ten percent, the total output of patents, i.e. the output of patents by all firms, would increase 20%, where the effect of spillovers accounts for more than half the increase. For the mean firm in the data, the coefficients and elasticities found by Jaffe imply a return of 2.0 patents per million dollars of own R&D expenditures, an 0.06 patents per million dollars of other firms' relevant R&D expenditures. For the models of profits and market value, he also finds interesting and statistically significant results for all elasticities and coefficients. The gross rate of return to R&D investment equals 27% and for capital it equals 15%, which is consistent with the idea that knowledge (R&D capital) has a higher rate of depreciation than (physical) capital. The effect the spillover pool has on profits and market value must be divided in a direct effect and an indirect effect. The direct effect is negative: the larger the spillover pool, i.e. the more R&D other firms perform, the lower both a firm's profits and its market value. However, the indirect effect which works through the interaction between a firm's total potential spillover pool and its own R&D efforts, is positive. For the firm with average logarithm of R&D effort, the net elasticity of profits with respect to the spillover pool is about 0.1. If all firms increased their R&D efforts by ten percent, aggregate profits, i.e. the profits of all the firms combined, would increase 3%, where the effect of spillovers accounts for about one-third of the increase. Similarly, the net elasticity of market value with respect to the spillover pool is about 0.05, evaluated at the average logarithm of R&D effort and the average logarithm of capital (p. 995). Finally, higher market shares lead to both higher profits and higher market value while a higher four-firm concentration ratio leads to both lower profits and lower market value for the average firm. Regarding the effect of technological opportunity on R&D output, profits and market value, the results are less clear. Although, the cluster dummies are all statistically significant, it is not certain that they actually result from differences in technological opportunity in the clusters. Using two different cross-sections corresponding to two different periods, however, Jaffe has found evidence that firms adjust their technological positions, i.e. the technological opportunities they face, in response to profit possibilities. The effect of these moves is that excess returns are dissipated.

Jaffe (1988)

Using much the same methodology as Jaffe (1986), Jaffe (1988) investigates both supply and demand factors that are expected to influence inventive activity, namely technological opportunity and R&D (knowledge) spillovers (supply), and market conditions (demand).

1. In the first step again the technological position, the proximity to other firms and the technological opportunity (21 technological clusters) of the firm is determined. However, now also the position of the firm in 'market space' is determined. In much the same way as its place in technology space, a firm's market position is described by how the firm's sales are spread over H separate and distinct markets. Hence, its market position can be denoted by a vector $G = (G_1, \ldots, G_H)$ where G_h is the fraction of the firm's sales achieved in market h (p. 432).
2. In the second step, Jaffe (1988) develops and tests descriptive equations of R&D intensity and productivity (TFP) growth. He models R&D intensity and productivity growth as being determined by technological position, industry effects (including market position) and the potential spillover pool. Using cross-sectional data from the NBER panel, Jaffe subsequently finds that spillovers are significant and that the positive externality outweighs any negative competition effects. When R&D expenditures in firms in close proximity to the firm increase, the firm increases its R&D expenditures as well. TFP subsequently grows faster even when the increase in the firm's own R&D efforts is controlled for. These results are consistent with the results of the earlier study by Jaffe (1986). In the R&D intensity equation more than half of the spillover effect results from firms outside the firm's technological cluster. In the TFP growth equation on the other hand, the potential spillover pool from outside the firm's technological cluster has no discernable effect on TFP growth. Furthermore, both technological position (technology push) and industry effects, e.g. market position (market pull), are found to be statistically significant in explaining R&D intensity, while they are only significant in explaining productivity growth when the potential spillover pool is not included in the estimating equation.

Megna and Mueller (1991)

Megna and Mueller try to explain why profit rates differ widely both across firms and industries. Their hypothesis is that these differences result from different intangible capital endowments, that is from differences in the R&D capital and the advertising capital firms have at their disposal. They subsequently estimate equations in which a firm's sales and a firm's profits respectively are a function of the firm's R&D capital, advertising capital but also of the R&D capital and the advertising capital of other firms in the same industry. Because these narrowly defined industries sell to final consumers and contain only a few firms (5-10), Megna and Mueller implicitly investigate interfirm pure knowledge spillovers.

They find that in their sample containing data on four industries (Games, toys and children's vehicles, Distilled liquor, Toilet preparations, and Pharmaceuticals), the effect of other firms' advertising capital generally leads to lower sales. The competitive effect of rivals' advertising thus dominates the positive externality of rivals' advertising increasing sales (and profits) for all firms selling similar products. For industry R&D, the results are more ambiguous, it can have both significant negative and significant positive effects on a firm's sales and profits. However, in pharmaceuticals, the most R&D intensive of the four industries examined, the positive spillover effect of rivals' R&D seems to dominate.

Acs, Audretsch and Feldman (1994)

Acs, Audretsch and Feldman investigate how receiving knowledge spillovers from R&D conducted elsewhere varies with firm size. They discern two possible sources of R&D spillovers, R&D performed in the firm's industry and research performed by universities, and two types of firms, large firms (> 500 employees) and small firms (< 500 employees). The dependent variable in the authors' regressions analysis is the number of innovations in a specific technological area and a (geographical) U.S. sate. Controlling for geographical variables like state population, they subsequently find that both types of spillovers contribute statistically significant to innovative output in both large and small firms. However, for large firms the R&D performed by other firms in the industry is the more important source while for small firms university research is the more important source of spillovers.

C.2 Intrafirm pure knowledge spillovers

Adams and Jaffe (1996)

Adams and Jaffe study both transfers of knowledge across different plants within a single firm and spillovers of knowledge across firms. In both cases, the extent to which knowledge's nonrivalrous nature will lead to increasing returns is moderated by other variables. The first moderating effect is exerted by technological distance (or inversely proximity). The idea is that the closer two firms are in technology space, the more useful the other's knowledge is to them, and similarly the larger the technological distance, the less useful the knowledge. A second moderating effect is that the benefits of knowledge's nonrivalrous can only be reaped effectively if the costs of the transfer are not prohibitive. That is, the costs associated with the transfer of knowledge, for example between different facilities of a single firm, may limit the extent of the increasing returns associated with the nonrivalrous nature of knowledge. Adams

and Jaffe construct estimating equations in which the effect of a firm's R&D efforts on productivity at the plant and at the firm level are affected by these and other variables. The variables they include are, amongst others, firm R&D effort, firm R&D capital, R&D effort per plant, the number of plants, the number of firm employees, and the relative plant age, i.e. the age of the plant compared to the average age of all the firm's plants. To account for the usefulness of interfirm spillovers they include a spillover pool and measures of both technological and geographical distance. Using data on firms in the chemicals industry, Adams and Jaffe start with estimating simple equations of plant productivity in which interfirm spillovers and distance effects are not taken into consideration (but the number of plants, relative plant age etc., are). The results are very interesting nonetheless. They find for instance that life-cycle effects are quite important. Measured productivity is lower both in new plants and in (older) plants that are cutting back. In addition, productivity increases with age, with each year (relative to the mean age of the firm's plants) leading to approximately 1% higher productivity. Another interesting finding is that firm R&D effort by itself does not have a statistically significant impact on plant productivity. It only becomes significant when the number of plants the firm comprises is taken into account. The number of plants in turn has a highly significant negative impact on productivity. In contrast with the absolute amount of R&D, which does not have a statistically significant impact on productivity, the amount of R&D per plant does have a highly statistically significant positive impact on productivity. The most interesting finding however, is that the impact of the number of plants is larger than that of R&D effort. That is, the spreading of R&D over different plants decreases the usefulness of R&D at such a rate that firm R&D efforts need to increase more than proportional to the number of plants, just to keep its overall contribution to plant productivity constant. This remarkable finding implies that plants would be better of as stand-alone facilities than as part of a larger establishment, and even that firms operating multiple plants would be better off if they were to break themselves up into independent firms. As a next step, Adams and Jaffe check the robustness of the striking result that the number of plants has a strong influence on the relation between R&D and plant productivity. The idea that the influence of the number of plants represents a firm-size effect can be dismissed instantly, however, because in the original regressions leading to the remarkable finding firm size was controlled for by taking the number of firm employees into account. Size was found to have a negative impact on the productivity level but loses its significance when the number of plants is included in the regression. The authors thus propose and test various other explanations. They start with relating levels of plant productivity to the number of plants a firm operates and the extent to

which a firm is centralized, i.e. whether the firm has a central head-office to which all plants report and whether it has one or more separate R&D facilities which supply knowledge to the manufacturing establishments. Adams and Jaffe find that the relation between the number of plants and productivity is hump-shaped (inverse U-shaped) and firms falling in the category of 11-20 plants are most productive. Similarly, they find that the relation between the number of plants and the centralization of the firm is U-shaped and firms falling in the category of 11-20 plants are most decentralized. Adams and Jaffe judge this pattern to be consistent with transfer pricing in which the most centralized firms transfer the value of plant output away from the manufacturing facilities towards the central head-office, thereby lowering measured output in these production plants. They study the issue further by investigating the effect the number of plants has on total output at the firm level. For this output is estimated using labor (number of employees), capital (total capital stock) and materials, as explanatory variables besides variables like R&D effort, R&D capital, R&D effort per plant and the number of plants. In this estimation, the number of plants is again found to have a statistically significant negative impact on productivity, output corrected for resource use. However, now R&D, that is R&D effort and R&D capital, has a statistically significant impact on plant productivity by itself, i.e. it is significant even when the number of plants is not controlled for. In fact, the coefficient for firm R&D is largely independent of whether the number of plants is included or not. Moreover, the coefficients for R&D and the number of plants are equal in magnitude which means that R&D does not have to increase more than proportional to an increase in the number of plants just to keep output corrected for resource use constant. Finally, both R&D effort per plant and R&D capital per plant are found to be significant in explaining productivity at the firm level. Based on these first findings, Adams and Jaffe draw two important conclusions regarding the influence of the number of plants on productivity. The first is that (i) the effect of the number of plants mainly works at the plant level and not at the firm level, indicating the importance of (mis-)measurement of plant output. The second is that (ii) dilution of knowledge (R&D) is important in countervailing the increasing returns knowledge's nonrivalrous nature gives rise to. The issue of R&D dilution, i.e. the diminishing contribution of firm R&D to plant level productivity as the number of plants a firm operates increases, is subsequently investigated further by including measures of both geographical and technological distance. The authors find that R&D performed far away geographically from the plant, i.e. outside the state where the plant is located or outside a 100 mile radius around the plant, is only between 10% and 30% as effective as R&D performed close by, i.e. within the same state as where the plant is located or within a 100 mile radius around the

plant. Even with the inclusion of measures of geographical distance, however, the dilution effect remains and the coefficients for the number of plants are generally larger than the coefficients for R&D. Adams and Jaffe further find that the effect of the number of plants is greater for the number of plants that are far away than it is for the number of plants that are close by. Technological distance has similar effects. The effectiveness of R&D performed outside the product field of the plant is only one-fourth to one-half as useful as R&D performed in the same product field. The dilution effect again remains and the elasticity of plant productivity with respect to the number of plants in the same product fields is larger than the elasticity with respect to firm R&D. The number of plants in other product fields contributes only little and not highly statistically significant to the dilution effect. However, the findings differ between a sample with data on firms in a range of industries and a sample with data on firms in the chemicals industry. Especially the relative effectiveness of R&D performed outside the product field of the plant differs between the samples. In the chemical industry sample, the effectiveness is only half as large as it is in the larger sample. Adams and Jaffe judge the effects of geographical and technological distance to be consistent with transfer pricing as well. Large numbers of far away plants indicate the existence of a central head-office located elsewhere. Similarly, the existence of a large number of plants active in the same product field indicates the existence of a central facility, e.g. a head-office, performing R&D for all these production units and skimming the value of their output. The authors continue with investigating the effects of both intrafirm and intraindustry pure knowledge spillovers and how their effect is conditioned by geographical distance. For this, they construct spillover pools, which consist of R&D performed by either other plants or other firms corrected for technological distance using the weighting scheme of Jaffe (1986). Adams and Jaffe find that the spillover pool constructed from R&D efforts by other firms in the industry corrected for technological distance contributes positively and statistically significant to plant productivity. Geographical proximity does not seem to matter to pure knowledge spillovers and the pool constructed from R&D performed more than 100 miles away seems to be as least as effective as the pool constructed from R&D performed within a 100 mile radius around the plant. What is more, the authors find that the elasticity of productivity with respect to the spillover pool is larger than the elasticity of productivity with respect to the firm's R&D efforts. The spillover effect from other firm's R&D is, however, also diluted by the number of plants in the industry. That is, plant productivity does not benefit from more R&D performed by other firms, but rather it benefits from other firms performing more R&D per plant. If subsequently the number of plants in the industry is not

included, the elasticity of productivity with respect to other firms' R&D efforts is diminished to one-third of the elasticity productivity has with respect to other firms' R&D efforts when the number of plants in the industry is included. Hence, productivity at the plant level is determined in part by other firms' R&D efforts per plant. Finally, Adams and Jaffe find that technological proximity and geographical proximity are related. Plants that are close-by are more likely to be in the similar product field than are firms that are more distant. The effect geography has on spillovers reported above may consequently be partly reflecting the effect technological distance has on spillovers. The most important conclusion Adams and Jaffe draw from their findings is that both interfirm and intrafirm spillovers are important but that the spillovers are diluted as the number of receivers increases. It is not so much the absolute amount of R&D that matters for productivity but rather the intensity. This of course is in conflict with the assumptions of the early models of R&D-based growth (e.g. Romer, Grossman and Helpman, and Aghion and Howitt). As possible explanations for their findings they propose (i) that there is always some technological distance which makes it impossible to use knowledge created by other plants and/or firms in exactly the same way as they do but that rather some adaptation is always required, and (ii) that internalizing knowledge spillovers beyond a certain point may run into increasing diseconomies of scale resulting from increasing organizational complexity and consequent loss of control (see for instance Williamson 1967).

Henderson and Cockburn (1996)

Henderson and Cockburn examine the relation between a firm's size and its research productivity. For this, they test four hypotheses using data for the period 1960-1988 on ten pharmaceutical firms, which account for 25% of all the research done in this industry. The hypotheses are, (1) individual research programs are not subject to returns to scale, (2) there are returns to scale in research at the firm level, (3) there are significant returns to scope in research, and (4) research productivity is positively associated with (pure) knowledge spillovers between firms (p. 35/36). Furthermore, the authors test three more hypotheses on how the importance of scale, scope and spillovers may change over time. The pharmaceutical industry provides a natural experiment to investigate whether the importance of scale, scope and spillovers changes over time because since 1978 the research process in the pharmaceutical industry has changed drastically. Research in this industry used to be a trial-and-error process in which potentially therapeutic substances were randomly tested for effectiveness. Little was known about the finer mechanism underlying the efficacy of a

certain substance. In 1978 however, for the first time, a wholesome substance was synthesized on the basis of deeper knowledge on the mechanism underlying its efficacy. Instead of looking for a therapeutic substance anywhere, research in the pharmaceutical industry become more directed on questions like finding a substance which stimulated the bonding of molecule A with molecule B, or inhibiting the breaking up of molecule C in molecule D and E, etc. This change in the conduct of pharmaceutical research may have changed the importance of scale, scope and spillover in this process. The three additional hypotheses subsequently are, (5) economies of scale in pharmaceutical research may have increased since 1978, (6) economies of scope in pharmaceutical research have increased since 1978, and (7) the impact of interfirm spillovers on research productivity has significantly increased since 1978 (p. 37). The authors start by analyzing firm-level data. In the estimating equation, they use output of patents as dependent variable and, amongst others, Size (total annual firm research expenditure), Research stock (accumulated research expenditure), Scope (the number of large research programs), Scope-squared ($Scope^2$) and Own-patent stock (the stock of the firm's own patents), as explanatory variables. Henderson and Cockburn find that all these variables have a positive and statistically significant relation to patent output at the firm level. Interestingly, the relation between patent output and Scope is hump-shaped (Scope - $Scope^2$) indicating that their is an optimal number of research programs for firms to be active in. Furthermore, firm dummies are most significant in explaining differences in patent output between the ten firms, which indicates that firm-specific factors like country of origin and the propensity to patent are important. Finally, the dummy variable indicating the period until 1978 (1960-1978), and after 1978 (1979-1988) is highly significant which suggests that there indeed has been a change in the way pharmaceutical research is conducted. The estimating equation the authors use to analyze program-level data contains some additional explanatory variables, besides Size, Research stock, Scope, and Own-patent stock. The extra explanatory variables are, Research (annual research expenditures on this program), Own-patent program stock (the stock of the firm's own patents in this program), Own-patents in related programs (intrafirm knowledge spillovers; flow of patents), Other-patents in the same program (interfirm knowledge spillovers; flow of patents), and Other patents in related programs (intrafirm spillovers; flow of patents). Size, Research stock (accumulated research expenditure), Scope and Own-patent stock are again found to have a positive and statistically significant relation to patent output at the program level with the relation between patent output and Scope being hump-shaped. For the other variables, Research, Own-patent program stock, Own-patents in related programs, Other-patents in the same program and Other patents

in related programs, much the same results are found. These variables all contribute positively and statistically significant to patent output. In addition, firm dummies and the period dummy (1960-1978/1979-1988) are again highly significant. Overall, Henderson and Cockburn find ample evidence supporting the first four of their hypotheses, that is, individual research programs are not subject to returns to scale, there are returns to scale in research at the firm level, there are significant returns to scope in research, and research productivity is positively associated with (pure) knowledge spillovers between firms. Regarding the final three hypotheses, the significance of the period dummy indicates that the nature of the research process has changed. In a direct comparison of the period 1960-1978 and 1979-1988 at the program level, Henderson and Cockburn confirm this suspicion. Almost all of the coefficients for the explanatory variables (e.g. Research, Size, and Scope) differ significantly over the two periods. Furthermore, the sixth hypothesis that economies of scope in pharmaceutical research have increased since 1978 is supported. About the other two hypotheses, that economies of scale in pharmaceutical research may have increased since 1978 and that the impact of interfirm spillovers on research productivity has increased significantly since 1978, no final judgement could be passed. The evidence however, indicates that since 1978 size is increasingly being used to exploit internal knowledge spillovers.

Klette (1996)

Klette investigates the relation between R&D and his measure of firm/plant (productivity) performance. He uses data for 1989 and 1990 on 804 Norwegian manufacturing plants to estimate equations in which performance is a function of, amongst others, performance in the previous year, the total amount of research done by the firm in the line of business to which the plant belongs, and dummy variables indicating the type of the plant (sole facility in a single-plant firm / parent facility in a multiplant firm / subsidiary facility in a multiplant firm). In a simple estimating equation, performance was shown to be positively - and highly statistically significant - related to previous year performance and the total amount of research. Most of the cross-plant variance in performance could be explained by previous year performance, indicating that performance differences may be persistent over time. The type of plant was also found to be significant in explaining differences in plant performance. In a more detailed estimating equation, Klette includes additional variables, besides past performance, the total amount of research and the type of plant, to take into account the effect of spillovers across plants active in the same line of business and belonging to the same group of firms. He finds that there are strong spillovers across firms within a group and a given line

of business (R&D performed outside the firm), especially when the R&D effort is large (the elasticity of performance with respect to R&D performed outside the firm is larger for larger R&D efforts). Klette concludes his study of the relation between R&D and (productivity) performance by investigating the effect of spillovers across lines of business both within the firm and within the group. There is a positive relation between performance and R&D effort in other lines of business within the firm but it is not statistically significant, which is probably due to the small sample size however. Similarly, there is a positive relation between performance and R&D effort in other lines of business within the group and this relation is statistically significant. For this last relation again the elasticity of performance with respect to the amount of R&D performed (in other lines of business within the group) is larger for larger R&D efforts.

Appendix D: The diversification-performance nexus

The first section of this Appendix Discusses in some detail some studies that summarize over 30 years of research on the diversification-performance nexus since the seminal contribution of Gort (1962) and 20 years since Rumelt's (1974) equally seminal study, namely Grant, Jammine and Thomas (1988), Ramanujam and Varadarajan (1989), Datta, Rajagopalan and Rasheed (1991), and Dess, Gupta, Hennart and Hill (1995). In the second section several recent studies of the relation between diversification and performance that seem to have taken at heart some of the harsh conclusions of these survey articles, namely Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), Lins and Servaes (1999), Rajan, Servaes and Zingales (2000), Campa and Kedia (2002) and Graham, Lemmon and Wolf (2002), are discussed.

D.1 Early work

Grant, Jammine and Thomas (1988)

Grant, Jammine and Thomas start their study of the causal relationships between diversity, diversification and profitability, with a discussion of the findings of previous research. They find that "… despite a burgeoning empirical literature, no consensus has emerged as to the impact of diversification on performance - indeed, recent studies have increased the inconsistencies between the findings of different researchers." (p. 771). They subsequently impute these inconsistencies to the fact that "Findings concerning relationships between diversity and profitability appear to be highly susceptible to choices concerning profitability measures, time period, control variables, and method of analysis." (p. 775). In addition, they state that "Although large-sample statistical research […] is a powerful means of identifying general relationships between pairs of variables, it is a comparatively weak method for gaining insight into complex interaction between variables that are not amenable to precise functional modeling." (p. 796). Grant, Jammine and Thomas thus conclude that, "To gain insight into the complex interaction between strategy, organization, and environment, detailed examination of the experiences of individual firms is needed." (p. 796). They subsequently propose replacing the coarse-grained research methods used previously with finer-grained research methods. Grant, Jammine and Thomas' review of the literature on the relation between diversification and performance can easily be summarized as follows. Grant,

Jammine and Thomas find that previous work investigating the diversification-performance nexus has provided disappointingly little results which are largely due to weaknesses in the research methods used in most work.

Ramanujam and Varadarajan (1989)

Ramanujam and Varadarajan express themselves on the research into the diversification-performance nexus in a similar fashion as Grant, Jammine and Thomas and reach much the same conclusions (as in fact do the other review articles discussed here). In an overall evaluation of the literature on diversification that follows their survey, they conclude that "The rationale for many of the studies […] seems to be nothing more ambitious than the opportunity to experiment with a different definition of diversification, a different measure of diversity, or a different approach to conceptualizing and measuring its effectiveness." (p. 536). They continue with stating that "It is difficult for us not to conclude […] that much rooms remains for breakthroughs in this seemingly oversaturated topic of inquiry." (p. 536/537). Ramanujam and Varadarajan attribute these disappointing results to five conceptual and methodological issues.

1. Empirical studies on the diversification-performance nexus have generally concentrated on only one or a few theoretical arguments for diversification and consequently used too narrow measures of performance.
2. Both the concept of diversification and performance are difficult to measure.
3. The cross-sectional nature of most research on the diversification-performance nexus which means that the results cannot be generalized over time and that possible important effects from the business cycle and the effects of diversity over larger time-frames (years) cannot be ascertained.
4. The lack of incorporation of variables other than diversity, e.g. market structure, R&D investment etc., which undoubtedly have a moderating effect on the relation between diversification and performance.
5. The fact that most research has examined the diversification-performance nexus by comparing the before and the after situation, and has had little attention for several issues surrounding the diversification process, e.g. implementation.

As a direction for future empirical research, Ramanujam and Varadarajan subsequently call for work that addresses these issues for instance through a comparative analysis of a small sample of diversification experiences.

Datta, Rajagopalan and Rasheed (1991)

Similar to Grant, Jammine and Thomas and Ramanujam and Varadarajan, Datta, Rajagopalan and Rasheed find that "… the findings of past empirical studies do not permit easy generalization - they have been, on the whole, inconclusive and often contradictory.' (p. 530) and that "… after almost 25 years of fairly intense research and investigation, very little can be said with certainty on the diversification-performance relationship." (p. 545). Datta, Rajagopalan and Rasheed divide their critique in theoretical issues and methodological issues. Under theoretical issues they share:

- the relative dominance of research into the link between degree of diversification (e.g. business count) and performance, type of diversification strategy (e.g. related) and performance, and mode of diversification (e.g. internal growth) and performance, respectively and the concomitant lack of incorporation of other moderating variables,
- the conceptualization of diversification as either degree of diversification, type of diversification, or mode of diversification, whilst the actual contribution of diversification to firm performance may lie in successfully (profitably) combining to a certain extent degree, type, and mode of diversification,
- the conceptualization of diversification as rationalized by economic objectives (e.g. profit) rather than other (non-economic) objectives such as managerial ones.

Under methodological issues Datta, Rajagopalan and Rasheed share:

- the uncritical pooling of cross-sectional data over multiple industries which leads to large sample heterogeneity,
- the difficulty of measuring degree and type of diversification, that is each method used to measure either degree or type of diversification has drawbacks in that it hides possible important information,
- the measurement of economic performance: any measure should include both risk and return in its operationalization and part of the diverging results regarding the performance consequences of diversification found by the articles surveyed can be attributed to the use of different performance measures.

In the overview article by Datta, Rajagopalan and Rasheed again the prevailing remark regarding future research is that it should take a different approach than hitherto in an attempt to address the theoretical and methodological issues raised (see also Grant, Jammine and Thomas and Ramanujam and Varadarajan).

Dess, Gupta, Hennart and Hill (1995)

Finally, Dess et al. conclude that "When all is said and done, we really do not have a very clear picture of the relationship between diversification strategy and performance. This despite twenty years of concentrated research effort on the part of scores of researchers, and dozens of published articles." (p. 370). Dess et al. formulate four broad points of critique, which unsurprisingly have a strong overlap to the critiques forwarded by Grant, Jammine and Thomas, Ramanujam and Varadarajan, and Datta, Rajagopalan and Rasheed.

1. The cross-sectional nature of most research on the diversification-performance nexus which makes it impossible to distinguish cause from effect and makes it harder to control for unobserved moderating factors.
2. The lack of attention for the issue of implementation and firms organizational structure (e.g. M-form).
3. The use made by most studies of the classification scheme developed in Rumelt's seminal work, which is flawed in that it lumps together the different economic benefits that might result from diversification.
4. The extreme focus of most empirical work on corporate strategy and corporate performance while diversification generally has its effects at the business unit level.

As fruitful directions for future research, Dess et al., propose, similar to the three overview articles discussed above, addressing these points of critique.

D.2 Recent work

Lang and Stulz (1994)

Lang and Stulz investigate the relation between the market's valuation of a firm, as measured by Tobin's q (cf. Tobin 1969), and its degree of diversification as measured by (i) a Herfindahl index computed from a firm's segment sales, (ii) a Herfindahl index computed from segment assets, and (iii) the total number of segments the firm comprises. As with most of the papers discussed here, the authors have taken the data used from the Compustat database, which contains data regarding U.S. firms. They find that in their sample covering the late 1970s and 1980s, diversified firms are valued lower in the capital markets than firms active in a single industry. In addition, they find that both the mean and median q of highly diversified firms, i.e. firms that report sales in five or more segments, is lower than both the mean and median of the whole sample for each year of the sample. The results cannot be explained on the ground of industry effects. Lang and Stulz compare portfolios of specialized

(single-segment) firms constructed to match the industry composition of diversified firms and still find a discount for diversified firms. This means that, on average, the shareholders of diversified firms would benefit from the breaking up of the diversified firms if this breaking up happened in such a way as to cause each newly independent segment to have the average q of the undiversified firms active in the same industry. Similarly, the results are robust when other explanatory variables (which are known to have a relation to q) such as size, access to capital markets and R&D intensity, are included. As Lang and Stulz (note, these results do not, however, mean that diversification hurts firm performance by definition. The performance discount might simply represent the fact that the individual segments of diversified firms have lower q than have their peers active in the same industry. In order to address this question of causality, the authors study the q of firms that diversify relative to firms that do not diversify. The results are that diversifying firms are poor performers relative to firms which do not diversify in general but that these firms do not perform significantly worse than do firms which are active in the same industry and do not diversify.

Berger and Ofek (1995)

Berger and Ofek start their study of the effect of diversification on firm value by calculating the value of diversified firms' segments as if they were operating as stand-alone firms. They impute this value by multiplying the median ratio of Capital (the sum of book value of debt and market value of equity) to Assets (or Sales or Earnings) for the corresponding single-segment firms active in the same industry by the segment's level of assets (or sales or earnings). The sum of these imputed values of a firm's segments is an estimate of what the value of the firm could/would have been had all its segments operated as stand-alone units. Berger and Ofek subsequently determine the effect of diversification by measuring excess value as the natural logarithm of the ratio of a firm's actual value to its imputed value. Positive excess value indicates that diversification enhances firm performance beyond that of operating its segments as stand-alone firms. The study finds that over the period 1986-1991 diversification, i.e. being active in more than one industry (four-digit SIC-code), leads to an average value loss between 13% and 15%. The discount is robust over the use of different ratio's used to calculate imputed values (Capital/Assets, Capital/Sales, or Capital/Earnings), is independent of firm size, and is lower when the segments of the firm involved are related, i.e. when the segments are in different three- and/or four-digit SIC-codes but in the same two-digit SIC-code. In addition, Berger and Ofek find that diversification reduces profitability. They subsequently investigate possible explanations for the performance effects of

diversification. A first possible explanation is offered by the finding that (i) overinvestment is associated with lower value and that diversified firms' segments overinvest more than stand-alone units do. Other evidence points to (ii) (cross-)subsidization of poorly performing segments as a second (partial) explanation for the value loss associated with diversification. These negative effects of diversification are mitigated by benefits of diversification like for instance tax advantages but Berger and Ofek estimate these to be limited, only 0.1% of sales.

Servaes (1996)

Servaes examines the diversification discount during the so-called 'conglomerate merger wave' of the late 1960s and early 1970s. Although Servaes uses the same data source as Lang and Stulz and Berger and Ofek did before him, he nonetheless comes up with several new and interesting results. First, he confirms (i) the existence of a diversification discount as indicated by lower value for diversified firms (a dummy variable taking value 1 when a firm is active in more than one industry (two-digit SIC-codes)) compared to single-segment firms during the 1960s, the time the merger wave started. He also finds, however, that (ii) the diversification discount declines over time, reaching a value of approximately zero in the mid-1970s. Servaes subsequently investigates whether industry effects and other variables like firm size, access to capital markets and R&D intensity (see also Lang and Stulz), can explain these results. However, neither of these factors was found able to account for either the level of the discount or its change over time. In addition, Servaes tests for causality and concludes that (iii) the results are not spurious: diversification leads to lower value (lower q) not the other way around. Finally, Servaes studies the effect the declining cost of diversification had on the decision to diversify and finds that (iv) the firms in his sample increased diversification the most in the period in which the discount was lowest, namely in the mid 1970s. This holds especially for firms with high insider ownership, which remained focused in the period 1961-1970 when the diversification discount was high, but were the first to diversify when the discount was low, i.e. in the period 1973-1976.

Lins and Servaes (1999)

The study of Lins and Servaes differs from the studies of Lang and Stulz, Berger and Ofek and Servaes, in that it examines international evidence on the diversification discount. They examine the effects of diversification for large samples of firms in Germany, Japan and the United Kingdom. Using the same measures of performance and diversification as Servaes, the authors find that the effect of diversification on performance differs across countries. After

adjusting for industry effects they find that in Germany the diversification discount equals zero while it equals 10% in Japan and 15% in the United Kingdom, which is very close to the value Berger and Ofek using much the same methodology found for the United States. Tests of robustness, however, reveal that the discount in Germany differs significantly from that in the United States only in 1992 and not in 1994 (the two years the sample comprises). The discount in Japan differs significantly from that in the United States for both years. Lins and Servaes subsequently investigate differences in ownership structure as a possible explanation for these differing discounts and find that concentrated insider ownership enhances the value of German firms but has no impact on the value of Japanese and U.K. firms. In Japan, however, the discounting effect of diversification only occurs for firms with strong ties to keiretsu groups.

Rajan, Servaes and Zingales (2000)

Rajan, Servaes and Zingales test the hypothesis that inefficient investment caused by internal power struggles gives rise to the diversification discount widely observed in empirical work on the relation between diversification and performance including the work discussed here. They find that within diversified firms segments with poor investment opportunities (as evidenced by low Tobin's q) receive more funds than these segments would have received as stand-alone business units. In addition, this inefficient investment increases with increasing diversity measured here as the dispersion of weighted opportunities. An increase in diversity in turn leads to lower firm value and there is thus strong reason to suspect inefficient investment as explaining (partly) the negative performance effects of diversification.

Campa and Kedia (2002)

Campa and Kedia take a closer look at the direction of causality in the relation between diversification and performance by investigating the (possible) endogeneity of a firm's choice to diversify. That is, they investigate whether firms are likely to diversify in response to circumstances which also have an independent effect on firm value. Campa and Kedia start with confirming the results of previous studies (including of course some of the ones discussed here) and find that diversification is associated with a significant discount. This of course is hardly surprising considering the fact that they basically use the same data and same methodology as these previous studies, i.e. the Compustat database, excess value as performance measure and a dummy indicating diversification (see for instance Berger and Ofek and Servaes). They control for the possible endogeneity in three ways, (1) by

introducing fixed-firm effects, (2) by modelling the decision to diversify as a function of characteristics of for instance the industry and the firm, and (3) by testing an endogenous self-selection model. When endogeneity is controlled for using either of these ways, the diversification discount drops and even turns into a premium, that is a negative discount, when the diversification decision is controlled for. Similar results are obtained in a sample of refocusing firms. The authors find that the decision to refocus is as much a function of variables that have an independent impact on performance, as diversification is.

Graham, Lemmon and Wolf (2002)

Graham, Lemmon and Wolf hypothesize that there are systematic differences between the segments of a diversified firm or conglomerate, and the stand-alone firms used as performance benchmarks. Not accurately accounting for these differences may incorrectly attribute negative (but also positive) performance to diversification where it should be attributed to these systematic differences. To investigate this issue of selection bias the authors use two samples of firms that have expanded the number of segments in which they are active. The first sample consists of (i) firms which have expanded their number of segments through mergers and/or acquisitions and for which the market value of both the acquiring firm and the target firm prior to the event could be identified. The second sample consists of (ii) firms that started of as single-segment firms but have expanded their number of segments (for instance through internal growth). The first sample allows for determining the effect of diversifying (through acquisition) and Graham, Lemmon and Wolf find an announcement return for the acquiring firm and the target firm combined of some 3%. Nonetheless, using the methodology of Berger and Ofek diversification through acquisition was found to lead to a 7% decline in excess value over the two-year period surrounding the event (t - 1, t + 1). There is a difference between types of acquisition however. For unrelated acquisitions (no common four-digit SIC codes) the change in excess value (-5%) is larger than for related acquisitions (-10.5%). Similarly, acquisitions which do not lead to an increase in the number of reported business segments lead to a smaller decline in excess value (-5.5%) than do acquisitions in which the number of reported segments does increase (-15%). Graham, Lemmon and Wolf find that almost all of the 7% decline in excess value concomitant to diversification through acquisition can be attributed to the acquiring of relatively poor performing targets. They find that the acquiring firms are valued at a premium compared to the typical stand-alone firms active in the same industries (excess value of 14%) while the target firms are poor performers relative to the typical single-segment firms active

in the same industries (excess value of -10%). Unsurprisingly, there is a value difference between targets in unrelated acquisitions (excess value of -20,5%) and related acquisitions (excess value of -4%). A similar result is found for targets in acquisitions that do not lead to an increased number of segments (excess value of -10.5%) and acquisition of targets which does lead to more business segments (excess value of -8%). Graham, Lemmon and Wolf use the second sample, that is firms which have expanded their number of segments through internal growth and/or reporting changes (one-third of the sample) or through mergers and/or acquisitions (two-thirds of the sample), to investigate how their findings for the first sample, that is firms which have expanded their number of segments through mergers and/or acquisitions, are related to the diversification discount reported by numerous studies including the ones discussed here. They find that the excess value of firms which increase their number of segments through acquisition declines with 14% while that of firms which increase their number of segments through internal growth and/or reporting changes shows no significant change. Graham, Lemmon and Wolf further find that in this second sample, again unsurprisingly, the excess value of target firms is significantly below zero in the year prior to their acquisitions and in fact shows a notable downward trend over the years prior to their acquisitions. Finally, they find that about one-half of the diversification discount of acquiring firms can be explained from the acquisition of an already discounted target firm.

Appendix E: Sectoral structure, value added per worker, and growth

Table E.1: Sectors.

Sector	ISIC group	Abbreviation
Agriculture	1	Ag
Mining	2	Mi
Manufacturing	3	Ma
Public utilities	4	PU
Construction	5	Co
Wholesale and retail trade	6	WaRT
Transport and communication	7	TaC
Finance, insurance and real estate	8	FIaRE
Community, social and personal service	9	CSaPS
Government services	10	GS
GDP at market prices	...	GDP

Remarks: Classification largely according to the International Standard Industrial Classification (ISIC) as used in van Ark (1996).

Table E.2: Sectoral shares in total output for the U.S. 1950-1990.

Year	Ag	Mi	Ma	PU	Co	WaRT	TaC	FIaRE	CSaPS	GS
1950	0.055	0.070	0.223	0.017	0.080	0.138	0.058	0.101	0.116	0.142
1951	0.049	0.071	0.227	0.017	0.081	0.127	0.059	0.097	0.108	0.164
1952	0.048	0.069	0.226	0.018	0.081	0.126	0.056	0.099	0.105	0.171
1953	0.048	0.069	0.234	0.019	0.081	0.126	0.055	0.101	0.104	0.164
1954	0.050	0.069	0.219	0.021	0.085	0.128	0.053	0.107	0.107	0.161
1955	0.048	0.071	0.227	0.020	0.085	0.131	0.054	0.108	0.106	0.150
1956	0.045	0.073	0.222	0.021	0.089	0.131	0.054	0.110	0.108	0.147
1957	0.043	0.070	0.217	0.022	0.086	0.129	0.053	0.113	0.110	0.145
1958	0.045	0.067	0.203	0.024	0.091	0.132	0.051	0.120	0.116	0.150
1959	0.041	0.066	0.213	0.024	0.093	0.133	0.051	0.120	0.115	0.142
1960	0.042	0.065	0.208	0.026	0.093	0.133	0.051	0.123	0.117	0.143
1961	0.040	0.065	0.205	0.026	0.092	0.131	0.050	0.126	0.119	0.146
1962	0.038	0.063	0.211	0.026	0.091	0.133	0.050	0.126	0.119	0.144
1963	0.037	0.063	0.218	0.027	0.090	0.132	0.050	0.125	0.119	0.141
1964	0.034	0.062	0.222	0.027	0.090	0.133	0.050	0.124	0.120	0.139
1965	0.033	0.060	0.228	0.027	0.088	0.134	0.051	0.124	0.118	0.136
1966	0.029	0.060	0.233	0.027	0.084	0.134	0.052	0.123	0.119	0.139
1967	0.030	0.061	0.227	0.028	0.080	0.134	0.052	0.125	0.121	0.143
1968	0.028	0.061	0.229	0.029	0.077	0.137	0.052	0.126	0.120	0.141
1969	0.028	0.061	0.228	0.030	0.072	0.135	0.053	0.130	0.122	0.140
1970	0.029	0.064	0.217	0.031	0.066	0.138	0.054	0.133	0.126	0.141
1971	0.029	0.062	0.216	0.032	0.063	0.142	0.054	0.137	0.126	0.138
1972	0.028	0.060	0.224	0.032	0.061	0.145	0.056	0.136	0.127	0.132
1973	0.026	0.056	0.234	0.034	0.059	0.145	0.056	0.135	0.128	0.126
1974	0.026	0.055	0.225	0.035	0.057	0.142	0.058	0.141	0.132	0.129
1975	0.028	0.054	0.211	0.037	0.053	0.147	0.057	0.145	0.136	0.133
1976	0.026	0.051	0.221	0.035	0.054	0.147	0.058	0.144	0.135	0.128
1977	0.025	0.050	0.227	0.034	0.054	0.148	0.059	0.143	0.137	0.124
1978	0.024	0.048	0.226	0.032	0.054	0.152	0.061	0.144	0.137	0.121
1979	0.024	0.045	0.226	0.030	0.053	0.155	0.062	0.146	0.139	0.119
1980	0.024	0.046	0.216	0.031	0.049	0.154	0.063	0.150	0.144	0.122
1981	0.027	0.046	0.213	0.031	0.047	0.157	0.061	0.150	0.146	0.121
1982	0.029	0.042	0.204	0.030	0.045	0.163	0.063	0.153	0.149	0.123
1983	0.024	0.040	0.209	0.029	0.045	0.164	0.067	0.153	0.149	0.120
1984	0.024	0.040	0.217	0.030	0.046	0.170	0.064	0.147	0.147	0.113

Year	Ag	Mi	Ma	PU	Co	WaRT	TaC	FIaRE	CSaPS	GS
1985	0.027	0.039	0.216	0.030	0.046	0.173	0.062	0.147	0.150	0.111
1986	0.028	0.034	0.216	0.030	0.047	0.178	0.062	0.144	0.152	0.110
1987	0.027	0.033	0.221	0.031	0.046	0.170	0.065	0.145	0.154	0.108
1988	0.025	0.036	0.223	0.031	0.044	0.171	0.065	0.146	0.153	0.105
1989	0.026	0.031	0.221	0.032	0.043	0.174	0.065	0.146	0.156	0.105
1990	0.027	0.034	0.217	0.032	0.042	0.169	0.067	0.145	0.159	0.107

Remarks: Output shares are calculated as sectoral Gross value added at market prices (in 1982 prices) divided by total Gross value added at market prices (in 1982 prices). Data is taken from van Ark (1996).

Table E.3: Linear trend in output shares in the U.S. 1950-1990.

Sector	Trend	Standard error
Agriculture	-6.85E-4	5.61E-5
Mining	-1.01E-3	4.53E-5
Manufacturing	-9.37E-5	1.06E-4
Public utilities	3.50E-4	4.05E-5
Construction	-1.40E-3	1.02E-4
Wholesale and retail trade	1.23E-3	7.69E-5
Transport and communication	3.43E-4	4.59E-5
Finance. insurance and real estate	1.29E-3	7.20E-5
Community. social and personal service	1.33E-3	4.70E-5
Government services	-1.33E-3	7.06E-5

Remarks: Trend rates (and standard errors) are calculated by estimating the following regression equation: Output share = $\beta_1 + \beta_2$*Year + ε with Year = the number of years elapsed since 1950. All trends rates are significant at a one percent critical level except that of manufacturing (p-value = 0.381).

Table E.4: Output per worker in the U.S. 1950-1990.

Year	Ag	Mi	Ma	PU	Co	WaRT	TaC	FIaRE	CSaPS	GS	Whole economy
1950	9631	87512	16984	35435	27578	14101	18779	57523	15809	19052	19059
1951	9942	95660	17623	40174	28630	13594	19614	57778	15741	20503	19729
1952	10460	98750	17897	42699	29971	13789	19501	59028	15916	20732	20146
1953	11139	106627	18320	45019	31453	14076	19728	59637	15928	20595	20571
1954	11412	114773	18221	49101	33810	14322	19746	60129	16262	20111	20861
1955	11959	126077	19491	50988	35148	15234	21185	62299	16108	19802	21725
1956	12341	126529	19280	54261	36195	15206	21472	63417	16097	19581	21890
1957	12545	126634	19495	57522	36939	15327	21669	65729	16400	19386	22452
1958	13384	131559	19266	58548	39466	15418	22093	66769	16515	19377	22402
1959	13287	144781	20531	64667	41302	16176	23240	69516	16795	19292	23273
1960	14332	149184	20426	68971	42474	16122	23707	71162	16833	19367	23509
1961	14393	158818	21027	72402	43272	16395	24528	72623	16883	19514	23957
1962	14730	167613	22068	76597	44255	17246	25679	74903	17150	19666	24717
1963	15479	179694	23649	80310	44540	17715	27170	76121	17532	19736	25577
1964	15927	187134	24894	85283	45215	18268	27775	77358	17972	19783	26287
1965	17000	192795	25909	87923	45358	18840	29547	79115	18193	19883	26986
1966	17213	203983	26211	92361	44661	19169	30947	80303	18353	19684	27195
1967	18645	218595	25834	94813	44201	19345	30576	80649	18601	19674	27262
1968	18318	228502	26696	102192	42722	19902	31710	80659	18650	19687	27673
1969	19364	232649	26884	106048	39420	19521	32557	82034	18828	19744	27752
1970	20824	241628	26495	106187	36543	19434	32711	81147	19134	19612	27673
1971	21739	242267	28169	111907	34740	20010	33379	82999	19187	19648	28280
1972	21488	242895	29927	113096	33838	20965	35774	83907	19436	19677	29028
1973	20850	235101	31379	124823	32161	21140	37070	82780	19788	19551	29423
1974	20190	210099	29907	126549	30649	20148	37244	83286	19749	19548	28741
1975	21609	188283	30243	134564	31400	20436	37225	83216	19589	19401	28797
1976	20946	177885	31948	133212	32579	20673	39699	84512	19758	19549	29375
1977	21719	171783	33053	131830	31616	20890	40885	83781	19956	19641	29702

Year	Ag	Mi	Ma	PU	Co	WaRT	TaC	FIaRE	CSaPS	GS	Whole economy
1978	21314	164230	33081	124745	30072	21325	42136	82627	19955	19671	29704
1979	22299	145697	33079	115995	28484	21561	42456	81582	19861	19128	29397
1980	21539	134363	32399	116667	26994	21173	42499	80456	19902	19024	28962
1981	24815	122850	32784	114665	27307	21668	41840	80071	19702	19209	29177
1982	25783	113215	32870	104482	27239	22066	43059	78341	19204	19140	28899
1983	21790	131611	35608	104068	27707	22586	48083	78839	19172	19176	29574
1984	24505	138176	37832	115889	27351	23751	47812	78443	19248	19117	30363
1985	29550	146779	39553	119475	26883	24299	47246	78232	19518	19149	30898
1986	32134	160250	41274	121630	27281	25294	48918	75810	19816	19209	31409
1987	31945	171371	43678	128942	27174	24311	51917	76024	19740	19196	31657
1988	30616	193975	44963	133691	26171	24813	52306	77946	19527	19381	32075
1989	31478	179217	45273	138030	26154	24959	52461	79285	19403	19441	31962
1990	34309	191341	45679	136892	25817	24624	53308	79008	19197	19542	31925

Remarks: Output per worker is calculated as Gross value added at market prices (in 1982 prices) divided by the Number of persons employed. Data is taken from van Ark (1996).

Table E.5: Growth in value added per worker in the U.S. 1950-1990.

Sector	Growth rate [%]	Standard error [%]	Growth rate larger than whole economy?
Agriculture	2.85	0.09	Yes
Mining	0.89	0.34	No
Manufacturing	2.43	0.06	Yes
Public utilities	2.99	0.22	No
Construction	-0.92	0.21	No
Wholesale and retail trade	1.52	0.05	Yes
Transport and communication	2.77	0.05	Yes
Finance, insurance and real estate	0.74	0.10	No
Community, social and personal service	0.63	0.05	No
Government services	-0.10	0.02	No
Whole economy	1.20	0.06	...

Remarks: Growth rates (and standard errors) are calculated by estimating the following regression equation: ln(Output per worker at market prices (in 1982 prices)) = $\beta_1 + \beta_2$*Year + ε with Year = the number of years elapsed since 1950. All growth rates are significant at a one percent critical level except for that of mining (p-value = 0.012).

Table E.6: Growth in value added in the U.S. 1950-1990.

Sector	Growth rate [%]	Standard error [%]	Growth rate larger than whole economy?
Agriculture	1.08	0.12	No
Mining	1.05	0.15	No
Manufacturing	2.97	0.09	No
Public utilities	4.39	0.19	Yes
Construction	0.83	0.17	No
Wholesale and retail trade	3.84	0.04	Yes
Transport and communication	3.60	0.07	Yes
Finance, insurance and real estate	4.04	0.08	Yes
Community, social and personal service	4.04	0.03	Yes
Government services	2.00	0.08	No
Whole economy	3.01	0.05	...

Remarks: Growth rates (and standard errors) are calculated by estimating the following regression equation: ln(Gross value added at market prices (in 1982 prices)) = $\beta_1 + \beta_2$*Year + ε with Year = the number of years elapsed since 1950. All growth rates are significant at a one percent critical level.

Appendix F: Growth, schooling, and income

Table F.1: Variables and definitions from Jones (1998b).

Variable	Definition
H	Average educational attainment in years (in 1995) (human capital)
Growth	Average annual growth rate of GDP per worker (between 1960 and 1997)
Y	GDP per worker relative to the U.S. (in 1997)
s_k	Average investment share of GDP

Table F.2: Variables and definitions.

Variable	Definition
Glow	Average annual growth rate of GDP per worker for lowest labor productivity-percentile
Glowmiddle	Average annual growth rate of GDP per worker for low-middle labor productivity-percentile
Gmiddlehigh	Average annual growth rate of GDP per worker for middle-high labor productivity-percentile
Ghigh	Average annual growth rate of GDP per worker for highest labor productivity-percentile

Remarks: Variables and definitions based on Jones (1998b).

Table F.3: Growth and human capital.

	Growth	Glow	Glowmiddle	Gmiddlehigh	Ghigh
H	0.00850 (4.04)***	0.00119 (0.565)	3.42E-4 (0.158)	0.00235 (1.28)	-0.00499 (3.24)***
H^2	-5.08E-4 (3.13)***	…	…	…	…
Number of countries	88	22	22	22	22
R^2 adj.	0.223	-0.0335	-0.0487	0.0294	0.311

Remarks: Absolute t-statistics in parentheses. (***) denotes significance at the one percent level. Data is taken from Jones (1998b).

Table F.4: Human capital and income.

	H	H
Y	8.17 (15.2)	14.04 (6.88)
Y^2	…	-6.76 (2.97)
Number of countries	88	88
R^2 adj.	0.726	0.749

Remarks: Absolute t-statistics in parentheses. All coefficients are significant at a one percent critical level. Data is taken from Jones (1998b).

Table F.5: Growth and income.

	Growth	s_k
Y	0.0921 (5.34)	0.367 (4.67)
Y^2	-0.0789 (3.98)	-0.209 (2.36)
Number of countries	109	97
R^2 adj.	0.305	0.490

Remarks: Absolute t-statistics in parentheses. All coefficients are significant at a one percent critical level except for that of Y^2 in explaining s_k (p-value = 0.0201). Data is taken from Jones (1998b).

Appendix G: Growth and sectoral structure

Table G.1: Growth rates of output per man-hour.

Country	1870-1950	1950-1973	1973-1987
Australia	0.0129	0.0266	0.0180
Austria	0.0135	0.0592	0.0275
Belgium	0.0132	0.0436	0.0319
Canada	0.0235	0.0295	0.0176
Finland	0.0201	0.0523	0.0221
France	0.0174	0.0498	0.0316
Germany	0.0149	0.0594	0.0259
Italy	0.0179	0.0578	0.0255
Netherlands	0.0133	0.0478	0.0238
Norway	0.0203	0.0424	0.0345
Sweden	0.0222	0.0440	0.0164
U.K.	0.0139	0.0319	0.0234
U.S.	0.0216	0.0246	0.0104

Remarks: Growth rates of output per man-hour are calculated as annual average compound growth rates. Data is taken from Maddison (1991).

Table G.2.A: Regression results.

	Growth	Growth	Growth	Growth
Share manufacturing	5.37E-4 (1.43)	6.28E-4 (1.14)	3.48E-4 (1.25)	2.57E-4 (0.602)
Country dummies exist?	No	Yes	No	Yes
Time dummies exist?	No	No	Yes	Yes
Degrees of freedom	37	25	35	23
Number of observations	39	39	39	39
R^2 adj.	0.0268	-0.240	0.645	0.675

Remarks: The U.S. functioned as the base country. Time dummies were included for the period 1870-1950 and 1950-1973. Share is employment share at the end of each period (i.e. 1950, 1973 and 1987). Absolute t-statistics in parentheses. Significance levels are denoted by (*) for ten percent, (**) for five percent, and (***) for one percent. Data is taken from Maddison (1991).

Table G.2.B: Regression results.

	Growth	Growth	Growth	Growth
Share manufacturing	5.67E-4 (1.98)*	1.09E-4 (2.82)***	1.38E-4 (0.596)	-7.06E-5 (0.177)
Country dummies exist?	No	Yes	No	Yes
Time dummies exist?	No	No	Yes	Yes
Degrees of freedom	37	25	35	23
Number of observations	39	39	39	39
R^2 adj.	0.0715	0.0104	0.633	0.670

Remarks: The U.S. functioned as the base country. Dummies were included for the period 1870-1950 and 1950-1973. Share is employment share at the beginning of each period (i.e. 1870, 1950 and 1973).Absolute t-statistics in parentheses. Significance levels are denoted by (*) for ten percent, (**) for five percent, and (***) for one percent. Data is taken from Maddison (1991).

Table G.3.A: Regression results.

	Growth	Growth	Growth	Growth
Share agriculture	2.33E-4 (1.15)	7.31E-4 (2.87)***	5.48E-4 (3.10)***	1.36E-4 (0.239)
Share manufacturing	5.29E-4 (1.41)	8.48E-4 (1.72)*	7.78E-4 (2.73)***	9.38E-5 (0.115)
Country dummies exist?	No	Yes	No	Yes
Time dummies exist?	No	No	Yes	Yes
Degrees of freedom	36	24	34	22
Number of observations	39	39	39	39
R^2 adj.	0.0352	0.0375	0.716	0.661

Remarks: The U.S. functioned as the base country. Time dummies were included for the period 1870-1950 and 1950-1973. Share is employment share at the end of each period (i.e. 1950, 1973 and 1987). Absolute t-statistics in parentheses. Significance levels are denoted by (*) for ten percent, (**) for five percent, and (***) for one percent. Data is taken from Maddison (1991).

Table G.3.B: Regression results.

	Growth	Growth	Growth	Growth
Share agriculture	2.92E-4 (1.58)	3.73E-4 (1.54)	7.31E-4 (4.81)***	8.21E-4 (2.30)**
Share manufacturing	1.17E-3 (2.47)**	2.02E-3 (2.84)***	8.23E-4 (3.05)***	7.40E-4 (1.46)
Country dummies exist?	No	Yes	No	Yes
Time dummies exist?	No	No	Yes	Yes
Degrees of freedom	36	24	34	22
Number of observations	39	39	39	39
R^2 adj.	0.108	0.0619	0.775	0.722

Remarks: The U.S. functioned as the base country. Dummies were included for the period 1870-1950 and 1950-1973. Share is employment share at the beginning of each period (i.e. 1870, 1950 and 1973).Absolute t-statistics in parentheses. Significance levels are denoted by (*) for ten percent, (**) for five percent, and (***) for one percent. Data is taken from Maddison (1991).

Appendix H: Growth scale effects in measurable sectors

Table H.1: Growth rates of output per worker measurable sectors period 1948-1992.

Country	Ag	Mi	Ma	PU	TaC
Denmark	0.0577	0.1398	0.0297	0.0662	0.0243
France	0.0547	0.0379	0.0439	0.0581	0.0403
Germany	0.0643	0.0175	0.0400	0.0441	0.0334
Italy	0.0543	0.0499	0.0512	0.0316	0.0339
Netherlands	0.0535	0.0748	0.0434	0.0519	0.0408
Spain	0.0536	0.0541	0.0574	0.0746	0.0508
Sweden	0.0429	0.0251	0.0345	0.0504	0.0369
U.K.	0.0477	0.0580	0.0293	0.0432	0.0293
U.S.	0.0323	0.0197	0.0250	0.0344	0.0277

Remarks: Output per worker is calculated as Gross value added at market prices (in 1982 prices) divided by the Number of persons employed. Growth rates of output per worker are calculated as annual average compound growth rates (from the earliest available data (i.e. 1948 or 1950) to the latest available data (i.e.1990 or 1992)). Data is taken from van Ark (1996).

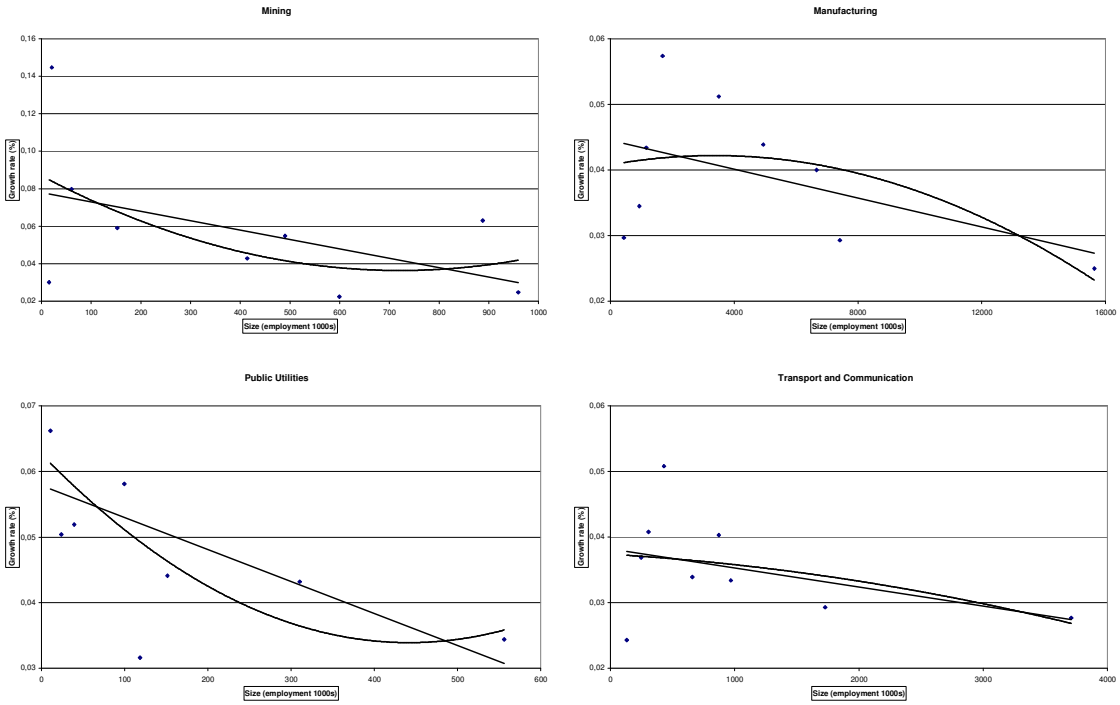


Figure H.1, Sector size and sectoral growth rate for the measurable sectors, Mining, Manufacturing, Public Utilities and Transport and Communication.
Source: van Ark (1996) and author's own calculations.

Table H.2: Sectoral growth of labor productivity and sector size.

	Growth	Growth
Size	-1.16E-6 (0.982)	-1.32E-6 (0.937)
Sectoral dummies included?	Yes	Yes
Country dummies included?	No	Yes
Degrees of freedom	39	31
Number of observations	45	45
R^2 adj.	0.044	0.181

Remarks: Size is employment size as measured at the begin of the period (i.e. 1948 or 1950). The Agricultural sector functioned as the base sector. When applicable the U.S. functioned as the base country. Absolute t-statistics in parentheses. Significance levels are denoted by (*) for five percent, and (**) for one percent. Absolute t-statistics in parentheses. All coefficients are significant at a one percent critical level except for that of Y^2 in explaining s_k (p-value = 0.020). Data is taken from van Ark (1996) and author's own calculations.

Appendix I: S-shaped vs. hump-shaped

This appendix shows the close relation between the S-shaped logistic curve and the hump-shaped pattern. Therefore the following equations are drawn in one figure:

1 $\text{Penetration}: P(t) = \dfrac{K}{1+e^{-(a+bt)}}$

2 $\text{Price}: \text{price}(t) = P_0 e^{-ct}$

3 $\text{Output}: P(t) \times \text{price}(t) = \dfrac{K}{1+e^{-(a+bt)}} \times P_0 e^{-ct}$

The idea is simply that when price declines faster than total output in physical units increases, total output measured in monetary units increases, and vice versa. Figure I.1 depicts the result.

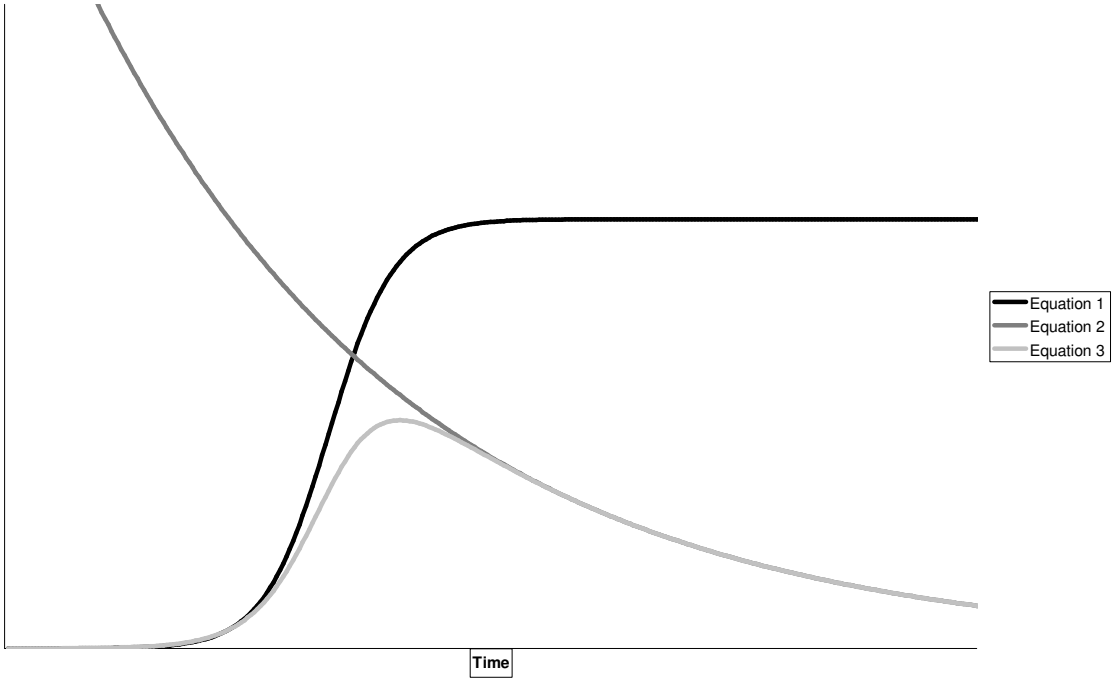


Figure I.1, S-shaped vs. hump-shaped.
Remarks: K = 1, a = 10, b = 1, P_0 = 2, and c = 0.1.

For output as a share subsequently much the same goes: when output of the product grows faster than total output, output as a share of total output increases and when output of the product grows slower than total output, output as a share of total output decreases.

References

Abramovitz, M. (1979), 'Economic Growth and Its Discontents', In: M.J. Boskin (ed.), *Economics and human welfare: essays in honor of Tibor Scitovsky*, New York, Academic Press, 3-21.

Abramovitz, M. (1993), 'The Search for the Sources of Growth: Areas of Ignorance, Old and New', *Journal of Economic History*, 53 (2), 217-43.

Acs, Z.J., D.B. Audretsch, and M.P. Feldman (1994), 'R&D spillovers and recipient firm size', *Review of Economics and Statistics*, 76 (2), 336-40.

Adams, J.D., and A.B. Jaffe (1996), 'Bounding the effects of R&D: an investigation using matched establishment-firm data', *RAND Journal of Economics*, 27 (4), 700-21.

Ades, A.F., and E.L. Glaeser (1999), 'Evidence on growth, increasing returns, and the extent of the market', *Quarterly Journal of Economics*, 114 (August), 1025-45.

Agarwal, R. (1998), 'Evolutionary trends of industry variables', *International Journal of Industrial Organization*, 16 (4), 511-25.

Agarwal, R., and M. Gort (1996), 'The evolution of markets and entry, exit and survival of firms', *Review of Economics and Statistics*, 78 (3), 489-98.

Aghion, P., and P. Howitt (1992), 'A model of growth through creative destruction', *Econometrica*, 60 (2), 323-51.

Aghion, P., and P. Howitt (1998), *Endogenous growth theory*, Cambridge MA, The MIT Press.

Aoki, M., and H. Yoshikawa (2002), 'Demand saturation-creation and economic growth', *Journal of Economic Behavior and Organization*, 48 (2), 127-54.

Ark, B. van (1996), 'Sectoral growth accounting and structural change in post-war Europe', In: B. van Ark and N. Crafts (eds.), *Quantitative aspects of post-war European economic growth*, Cambridge, Cambridge University Press, 84-164.

Arrow, K. (1962), 'Economic Welfare and the Allocation of Resources for Invention', In: R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, Princeton University Press, 609-25.

Aw, B.-Y. (2002), 'Productivity Dynamics of Small and Medium Enterprises in Taiwan', *Small Business Economics*, 18 (1-3), 69-84.

Backus, D.K., P.J. Kehoe, and T.J. Kehoe (1992), 'In Search of Scale Effects in Trade and Growth', *Journal of Economic Theory*, 58 (2), 377-409.

Bahk, B.-H., and M. Gort (1993), 'Decomposing Learning by Doing in New Plants', *Journal of Political Economy*, 101 (4), 561-83.

Baily, M.N., and R.J. Gordon (1988), 'The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power', *Brookings Papers on Economic Activity,* 2, 347-420.

Barro R.J., and X. Sala-i-Martin (1995), *Economic Growth*, New York, McGraw-Hill.

Baumol, W.J. (1967), 'Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis', *American Economic Review*, 57 (3), 415-26.

Baumol, W.J., S.A. Batey Blackman, and E.N. Wolff (1989), *Productivity and American Leadership*, Cambridge MA, The MIT Press.

Berger, P.G., and E. Ofek (1995), 'Diversification's effect on firm value', *Journal of Financial Economics*, 37, 39-65.

Bils, M., and P.J. Klenow (2001), 'The Acceleration in Variety Growth', *American Economic Review***,** 91 (2), 274-80.

Blanchflower, D.G., and A.J. Oswald (2003), 'Well-being over time in Britain and the USA', *Journal of Public Economics*, in press.

Bloom, N., and J. van Reenen (2002), 'Patents, real options and firm performance', *Economic Journal*, 112 (March), C97-116.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches, and D.W. Jorgenson (1996), *Toward A More Accurate Measure Of The Cost Of Living*, the final report from the 'Boskin Commission' to the Senate Finance Committee, available at http://www.ssa.gov/history/reports/boskinrpt.html.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches, and D.W. Jorgenson (1997), 'The CPI Commission: Findings and Recommendations', *American Economic Review, Papers and Proceedings*, 87 (2), 78-83.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches, and D.W. Jorgenson (1998), 'Consumer Prices, the Consumer Price Index, and the Cost of Living', *Journal of Economic Perspectives*, 12 (1), 3-26.

Boskin, M.J., and D.W. Jorgenson (1997), 'Implications of Overstating Inflation for Indexing Government Programs and Understanding Economic Progress', *American Economic Review, Papers and Proceedings*, 87 (2), 89-93.

Bosworth, D.L. (1978), 'The Rate of Obsolescence of Technical Knowledge - A Note', *Journal of Industrial Economics*, 26 (3), 273-79.

Bound, J., C. Cummins, Z. Griliches, B.H. Hall, and A. Jaffe (1984), 'Who Does R&D and Who Patents?', In: Z. Griliches (ed.), *R&D, Patents, and Productivity*, Chicago, The University of Chicago Press, 21-54.

Caballero, R.J., and A.B. Jaffe (1993), 'How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth', In: O.J. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual 1993*, Cambridge MA, The MIT Press, 15-74.

Campa, J.M., and S. Kedia (2002), 'Explaining the Diversification Discount', *Journal of Finance*, 57 (4), 1731-62.

Clark, C. (1951), *The conditions of economic progress*, second rewritten edition (1940), London, MacMillan.

Coe, D.T., and E. Helpman (1995), 'International R&D spillovers', *European Economic Review*, 39, 859-87.

Coe, D.T., E. Helpman, and A.W. Hoffmaister (1997), 'North-south R&D spillovers', *Economic Journal*, 107 (January), 134-49.

Cohen, W.M., and S. Klepper (1996a), 'Firm size and the nature of innovation within industries: the case of process and product innovation', *Review of Economics and Statistics*, 78 (2), 232-43.

Cohen, W.M., and S. Klepper (1996b), 'A reprise of size and R&D', *Economic Journal*, 106 (July), 925-51.

Cohen, W.M., R.C. Levin, and D.C. Mowery (1987), 'Firm size and R&D intensity: a re-examination', *Journal of Industrial Economics*, 35 (4), 543-65.

Cohen, W.M., and D.A. Levinthal (1989), 'Innovation and learning: the two faces of R&D', *Economic Journal*, 99 (September), 569-96.

Datta, D.K., N. Rajagopalan, and A.M.A. Rasheed (1991), 'Diversification and performance: critical review and future directions', *Journal of Management Studies*, 28 (5), 529-58.

Deaton, A., and J. Muellbauer (1983), *Economics and consumer behavior*, reprinted with corrections (1980), Cambridge, Cambridge University Press.

Dess, G.G., A. Gupta, J.-F. Hennart, and C.W.L. Hill (1995), 'Conducting and Integrating Strategy Research at the International, Corporate, and Business Levels: Issues and Directions', *Journal of Management*, 21 (3), 357-93.

Dinopoulos, E., and P. Thompson (1998), 'Schumpeterian Growth without Scale Effects', *Journal of Economic Growth*, 3 (4), 313-35.

Dixit, A.K., and J.E. Stiglitz (1977), 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review*, 67 (3), 297-308.

Dixon, R. (1980), 'Hybrid corn revisited', *Econometrica*, 48 (6), 1451-61.

Doms, M., T. Dunne, and M.J. Roberts (1995), 'The role of technology use in the survival and growth of manufacturing plants', *International Journal of Industrial Organization*, 13 (4), 523-42.

Dosi, G. (1988), 'Sources, Procedures, and Microeconomic Effects of Innovation', *Journal of Economic Literature*, 26 (September), 1126-71.

Duesenberry, J.S. (1949), *Income, saving and the theory of consumer behavior*, Cambridge MA, Harvard University Press.

Dunne, P., and A. Hughes (1994), 'Age, size, growth and survival: UK companies in the 1980s', *Journal of Industrial Economics*, 42 (2), 115-40.

Dunne, T. (1994), 'Plant age and technology use in U.S. manufacturing industries', *RAND Journal of Economics*, 25 (3), 488-99.

Dunne, T., M.J. Roberts, and L. Samuelson (1988), 'Patterns of firm entry and exit in U.S. manufacturing industries', *RAND Journal of Economics*, 19 (4), 495-515.

Dunne, T., M.J. Roberts, and L. Samuelson (1989), 'The growth and failure of U.S. manufacturing plants', *Quarterly Journal of Economics*, 104 (November), 671-98.

Easterlin, R.A. (1974), 'Does Economic Growth Improve the Human Lot? Some Empirical Evidence', In: P.A. David and M.W. Reder (eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, New York and London, Academic Press, 89-125.

Easterlin, R.A. (1995), 'Will raising the incomes of all increase the happiness of all?', *Journal of Economic Behavior and Organization*, 27 (1), 35-47.

Easterlin, R.A. (2001), 'Income and happiness: towards a unified theory', *Economic Journal*, 111 (July), 465-84.

Echevarria, C. (1997), 'Changes in sectoral composition associated with economic growth', *International Economic Review*, 38 (2), 431-52.

Evans, D.S. (1987a), 'The relationship between firm growth, size, and age: estimates for 100 manufacturing industries', *Journal of Industrial Economics*, 35 (4), 567-81.

Evans, D.S. (1987b), 'Tests of Alternative Theories of Firm Growth', *Journal of Political Economy*, 95 (4), 657-74.

Falkinger, J., and J. Zweimüller (1996), 'The cross-country Engel curve for product diversification', *Structural Change and Economic Dynamics*, 7 (1), 79-97.

Falkinger, J., and J. Zweimüller (1997), 'The impact of income inequality on product diversity and economic growth', *Metroeconomica*, 48 (3), 211-37.

Fisher, F.M., and P. Temin (1973), 'Returns to Scale in Research and Development: What Does the Schumpeterian Hypothesis Imply?', *Journal of Political Economy*, 81 (1), 56-70.

Föllmi, R., and J. Zweimüller (2000), *Structural Change and Economic Growth with A Hierarchy of Needs*, Working Paper, Zurich, University of Zurich.

Föllmi, R., and J. Zweimüller (2002), *Structural Change and the Kaldor Facts of Economic Growth*, Working Paper, Zurich, University of Zurich.

Frank, R.H. (1999), *Luxury fever*, New York, The Free Press.

Frantzen, D. (2002), 'R&D Spillovers and Scale Effects', *Jahrbücher für Nationalökonomie und Statistik*, 222 (6), 727-49.

Furman, J.L., M.E. Porter, and S. Stern (2002), 'The determinants of national innovative capacity', *Research Policy*, 31, 899-933.

Gordon, R.J. (2000), 'Does the "New Economy" Measure up to the Great Inventions of the Past?', *Journal of Economic Perspectives*, 14 (4), 49-74.

Gordon, R.J., and Z. Griliches (1997), 'Quality Change and New Products', *American Economic Review, Papers and Proceedings*, 87 (2), 84-8.

Gort, M. (1969), 'An economic disturbance theory of mergers', *Quarterly Journal of Economics*, 83 (November), 624-42.

Gort, M., and S. Klepper (1982), 'Time paths in the diffusion of product innovations', *Economic Journal*, 92 (September), 630-53.

Gort, M. (1962), *Diversification and Integration in American Industry*, Princeton, Princeton University Press.

Goto, A., and K. Suzuki (1989), 'R&D capital, rate of return on R&D investment and spillover of R&D in Japanese manufacturing industries', *Review of Economics and Statistics*, 71 (4), 555-64.

Grabowski, H.G. (1968), 'The Determinants of Industrial Research and Development: A Study of the Chemical, Drug, and Petroleum Industries', *Journal of Political Economy*, 76 (2), 292-306.

Graham, J., M. Lemmon, and J. Wolf (2002), 'Does Corporate Diversification Destroy Value?', *Journal of Finance*, 57 (2), 695-720.

Grant, R.M., A.P. Jammine, and H. Thomas (1988), 'Diversity, diversification, and profitability among British manufacturing companies, 1972-84', *Academy of Management Journal*, 31 (4), 771-801.

Griliches, Z. (1957), 'Hybrid corn: an exploration in the economics of technological change', *Econometrica*, 25 (4), 501-22.

Griliches, Z. (1961), 'Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change', In: The Price Statistics of the Federal Government, General Series, #73, New York, NBER, 137-96. Reprinted in: Griliches, Z., ed. (1971), *Price Indexes and Quality Change*, Cambridge MA, Harvard University Press, 55-87.

Griliches, Z. (1971a), 'Introduction: Hedonic Price Indexes Revisited', In: Z. Griliches (ed.), *Price Indexes and Quality Change*, Cambridge MA, Harvard University Press, 3-15.

Griliches, Z. (1979), 'Issues in assessing the contribution of research and development to productivity growth', *Bell Journal of Economics*, 10 (1), 92-116.

Griliches, Z. (1990), 'Patent Statistics as Economic Indicators: A Survey', *Journal of Economic Literature*, 28 (December), 1661-1707.

Griliches, Z. (1992), 'The Search for R&D Spillovers', *Scandinavian Journal of Economics*, 94, S29-47.

Griliches, Z. (1994), 'Productivity, R&D, and the Data Constraint', *American Economic Review*, 84 (1), 1-23.

Griliches, Z. (1996), 'The Discovery of the Residual: A Historical Note', *Journal of Economic Literature*, 34 (September), 1324-30.

Griliches, Z., and F.R. Lichtenberg (1984), 'R&D and Productivity Growth at the Industry Level: Is There Still a Relationship', In: Z. Griliches (ed.), *R&D, Patents, and Productivity*, Chicago, The University of Chicago Press, 465-96.

Griliches, Z., ed. (1971b), *Price Indexes and Quality Change*, Cambridge MA, Harvard University Press.

Griliches, Z., ed. (1984), *R&D, Patents, and Productivity*, Chicago, The University of Chicago Press.

Grossman, G.M., and E. Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge MA, The MIT Press.

Hall, B.H. (1987), 'The relationship between firm size and firm growth in the U.S. manufacturing sector', *Journal of Industrial Economics*, 35 (4), 583-606.

Henderson, R., and I. Cockburn (1996), 'Scale, scope, and spillovers: the determinants of research productivity in drug discovery', *RAND Journal of Economics*, 27 (1), 32-59.

Hirsch, F. (1976), *The Social Limits to Growth*, Cambridge MA, Harvard University Press.

Hirsch, S. (1967), *Location of industry and international competitiveness*, Oxford, Clarendon Press.

Houthakker, H.S. (1987a), 'Engel, Ernst', In: J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Vol. 2, London, MacMillan, 142.

Houthakker, H.S. (1987b), 'Engel's Law', In: J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Vol. 2, London, MacMillan, 143-4.

Idson, T.L., and W. Oi (1999), 'Workers Are More Productive in Large Firms', *American Economic Review, Papers and Proceedings*, 89 (2), 104-8.

Illueca, M., and J.A. Lafuente (2003), 'Productivity and Scale Effect in Closely Related Firms: Evidence from the Spanish Tile Sector', *International Small Business Journal*, 21 (2), 161-80.

Jaffe, A.B. (1986), 'Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value', *American Economic Review*, 76 (5), 984-1001.

Jaffe, A.B. (1988), 'Demand and supply influences in R&D intensity and productivity growth', *Review of Economics and Statistics*, 70 (3), 431-7.

Jensen, M.C. (1986), 'Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers', *American Economic Review, Papers and Proceedings*, 76 (2), 323-9.

Jones, C.I. (1995a), 'Time series tests of endogenous growth models', *Quarterly Journal of Economics*, 110 (May), 495-525.

Jones, C.I. (1995b), 'R&D-Based Models of Economic Growth', *Journal of Political Economy*, 103 (4), 759-84.

Jones, C.I. (1999), 'Growth: With or Without Scale Effects?', *American Economic Review, Papers and Proceedings*, 89 (2), 139-44.

Jones, C.I. (1998a), *Introduction to economic growth*, New York, Norton.

Jones, C.I. (1998b), *Introduction to economic growth*, second edition 2002, New York, Norton.

Jovanovic, B. (1982), 'Selection and the evolution of industry', *Econometrica*, 50 (3), 649-70.

Jovanovic, B., and S. Lach (1989), 'Entry, Exit, and Diffusion with Learning by Doing', *American Economic Review*, 79 (2), 690-9.

Jovanovic, B., and G.M. MacDonald (1994a), 'Competitive Diffusion', *Journal of Political Economy*, 102 (1), 24-52.

Jovanovic, B., and G.M. MacDonald (1994b), 'The Life Cycle of a Competitive Industry', *Journal of Political Economy*, 102 (2), 322-47.

Kamien, M.I., and N.L. Schwartz (1982), *Market structure and innovation*, Cambridge, Cambridge University Press.

Kleinknecht, A., and B. Verspagen (1990), 'Demand and innovation: Schmookler re-examined', *Research Policy*, 19, 387-94.

Klepper, S. (1996), 'Entry, Exit, Growth and Innovation over the Product Cycle', *American Economic Review*, 86 (3), 562-83.

Klepper, S., and E. Graddy (1990), 'The evolution of new industries and the determinants of market structure', *RAND Journal of Economics*, 21 (1), 27-44.

Klette, T.J. (1996), 'R&D, scope economies and plant performance', *RAND Journal of Economics*, 27 (3), 502-22.

Kongsamut, P., S. Rebelo, and D. Xie (2001), 'Beyond Balanced Growth', *Review of Economic Studies*, 68, 869-82.

Kortum, S. (1997), 'Research, Patenting and Technological Change', *Econometrica*, 65 (6), 1389-1419.

Kremer, M. (1993), 'Population growth and technological change: one million B.C. to 1990', *Quarterly Journal of Economics*, 108 (August), 681-716.

Kumar, M.S. (1985), 'Growth, acquisition activity and firm size: evidence from the United Kingdom', *Journal of Industrial Economics*, 33 (3), 327-38.

Lach, S. (1995), 'Patents and productivity growth at the industry level: a first look', *Economics Letters*, 49, 101-8.

Laitner, J. (2000), 'Structural Change and Economic Growth', *Review of Economic Studies*, 67, 545-61.

Lang, L., and R. Stulz (1994), 'Tobin's q, Corporate Diversification, and Firm Performance', *Journal of Political Economy*, 102 (6), 1248-80.

Layard, R. (1980), 'Human satisfactions and public policy', *Economic Journal*, 90 (December), 737-50.

Leibenstein, H. (1950), 'Bandwagon, snob, and Veblen effects in the theory of consumer demand', *Quarterly Journal of Economics*, 64 (May), 183-207.

Lemelin, A. (1982), 'Relatedness in the patterns of interindustry diversification', *Review of Economics and Statistics*, 64 (4), 646-57.

Lewellen, W.G. (1971), 'A pure financial rationale for the conglomerate merger', *Journal of Finance*, 26 (2), 521-37.

Li, C.-W. (2000), 'Endogenous vs. semi-endogenous growth in a two-R&D-sector model', *Economic Journal*, 110 (March), C109-22.

Li, C.-W. (2002), 'Growth and scale effects: the role of knowledge spillovers', *Economics Letters*, 74, 177-85.

Lichtenberg, F.R. (1992), 'Industrial de-diversification and its consequences for productivity', *Journal of Economic Behavior and Organization*, 18 (3), 427-38.

Link, A.N. (1980), 'Firm Size and Efficient Entrepreneurial Activity: A Reformulation of the Schumpeter Hypothesis', *Journal of Political Economy*, 88 (4), 771-82.

Link, A.N., and J.E. Long (1981), 'The simple economics of basic scientific research: a test of Nelson's diversification hypothesis', *Journal of Industrial Economics*, 30 (1), 105-9.

Lins, K., and H. Servaes (1999), 'International Evidence on the Value of Corporate Diversification', *Journal of Finance*, 54 (6), 2215-39.

Lucas, R.E. Jr. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, 22 (1), 3-42.

Lucas, R.E. Jr. (1993), 'Making a miracle', *Econometrica*, 61 (2), 251-72.

MacDonald, J.M. (1985), 'R&D and the directions of diversification', *Review of Economics and Statistics*, 67 (4), 583-90.

Machlup, F. (1980), *Knowledge and Knowledge Production*, Vol. I of Knowledge: its creation, distribution and economic significance, Princeton, Princeton University Press.

Maddison, A. (1991), *Dynamic forces in capitalist development: a long-run comparative view*, Oxford, Oxford University Press.

Majumdar, S.K. (1997), 'The Impact of Size and Age on Firm-Level Performance: Some Evidence from India', *Review of Industrial Organization*, 12 (2), 231-41.

Mankiw, N.G. (1997), *Principles of Economics*, Forth Worth, The Dryden Press.

Mansfield, E. (1962), 'Entry, Gibrat's Law, Innovation, and the Growth of Firms', *American Economic Review*, 62 (5), 1023-51.

Maslow, A.H. (1970), *Motivation and Personality*, second revised edition (1954), New York, Harper & Row.

Matsusaka, J.G. (2001), 'Corporate Diversification, Value Maximization, and Organizational Capabilities', *Journal of Business*, 74 (3), 409-31.

Matsuyama, K. (2000), 'A Ricardian Model with a Continuum of Goods under Nonhomothetic Preferences: Demand Complementarities, Income Distribution, and North-South Trade', *Journal of Political Economy*, 108 (6), 1093-1120.

Matsuyama, K. (2002), 'The Rise of Mass Consumption Societies', *Journal of Political Economy*, 110 (5), 1035-70.

McBride, M. (2001), 'Relative-income effects on subjective well-being in the cross-section', *Journal of Economic Behavior and Organization*, 45 (3), 251-78.

McNulty, P.J. (1974), 'On Firm Size and Innovation in the Schumpeterian System', *Journal of Economic Issues*, 8 (3), 627-32.

Megna, P., and D.C. Mueller (1991), 'Profit rates and intangible capital', *Review of Economics and Statistics*, 73 (4), 632-42.

Montgomery, C.A. (1994), 'Corporate Diversification', *Journal of Economic Perspectives*, 8 (3), 163-78.

Montgomery, C.A., and S. Hariharan (1991), 'Diversified expansion by large established firms', *Journal of Economic Behavior and Organization*, 15 (1), 71-89.

Mueller, D.C. (1969), 'A theory of conglomerate mergers', *Quarterly Journal of Economics*, 83 (November), 643-59.

Murphy, K.M., A. Shleifer, and R. Vishny (1989), 'Income distribution, market size, and industrialization', *Quarterly Journal of Economics*, 104 (August), 537-64.

Nelson, R.A. (1990), 'Productivity Growth, Scale Economies and the Schumpeterian Hypothesis', *Southern Economic Journal*, 57 (2), 521-7.

Nelson, R.R. (1959), 'The simple economics of basic scientific research', *Journal of Political Economy*, 67 (3), 297-306.

Nelson, R.R., ed. (1962), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, Princeton University Press.

Nelson, R.R., M.J. Peck, and E.D. Kalachek (1967), *Technology, Economic Growth and Public Policy*, Washington, The Brookings Institution.

Nelson, R.R., and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge MA, Harvard University Press.

Nickell, S.J. (1996), 'Competition and Corporate Performance', *Journal of Political Economy*, 104 (4), 724-46.

Nordhaus, W.D. (1969), *Invention, Growth, and Welfare*, Cambridge MA, The MIT Press.

Oswald, A.J. (1997), 'Happiness and economic performance', *Economic Journal*, 107 (September), 1815-31.

Pagano, P, and F. Schivardi (2003), 'Firm Size Distribution and Growth', *Scandinavian Journal of Economics*, 105 (2), 255-74.

Panzar, J.C., and R.D. Willig (1981), 'Economies of Scope', *American Economic Review, Papers and Proceedings*, 71 (2), 268-72.

Papageorgiou, C. (2002), 'On scale effects: further evidence', *Applied Economics Letters*, 9, 245-9.

Pakes, A., and Z. Griliches (1984), 'Patents and R&D at the Firm Level: A First Look', In: Z. Griliches (ed.), *R&D, Patents, and Productivity*, Chicago, The University of Chicago Press, 55-72.

Pasinetti, L.L. (1981), *Structural change and economic growth*, Cambridge, Cambridge University Press.

Penrose, E.T. (1959), *The Theory of the Growth of the Firm*, third edition 1995, Oxford, Basil Blackwell.

Peretto, P.F. (1998), 'Technological Change and Population Growth', *Journal of Economic Growth*, 3 (4), 283-311.

Peretto, P.F., and S. Smulders (2002), 'Technological distance, growth and scale effects', *Economic Journal*, 112 (July), 603-24.

Pigou, A.C. (1952), *The economics of welfare*, fourth edition (1920), reprinted with eight new appendices 1952, London, MacMillan.

Porter, M.E., and S. Stern (2000), *Measuring the "Ideas" Production Function: Evidence from International Patent Output*, NBER Working Paper #7891, Cambridge MA, NBER.

Rajan, R., H. Servaes, and L. Zingales (2000), 'The Cost of Diversity: The Diversification Discount and Inefficient Investment', *Journal of Finance*, 55 (1), 35-80.

Rapping, L. (1965), 'Learning and World War II production functions', *Review of Economics and Statistics*, 47 (1), 81-6.

Ramanujam, V., and P. Varadarajan (1989), 'Research on corporate diversification: a synthesis', *Strategic Management Journal*, 10, 523-51.

Romer, P.M. (1986), 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, 94 (5), 1002-37.

Romer, P.M. (1990), 'Endogenous Technological Change', *Journal of Political Economy*, 98 (5), S71-102.

Rotemberg, J.J., and G. Saloner (1994), 'Benefits of Narrow Business Strategies', *American Economic Review*, 84 (5), 1330-49.

Rowthorn, R., and R. Ramaswamy (1999), 'Growth, Trade, and Deindustrialization', *IMF Staff Papers*, 46 (1), 18-41.

Rumelt, R.P. (1974), *Strategy, Structure and Economic Performance*, Harvard Business School Classic 1986, Boston, Harvard Business School Press.

Sala-i-Martin, X. (1997), 'I Just Ran Two Million Regressions', *American Economic Review, Papers and Proceedings*, 87 (2), 178-83.

Scherer, F.M. (1965), 'Firm Size, Market Structure, Opportunity and the Output of Patented Inventions', *American Economic Review*, 55 (5), 1097-125.

Scherer, F.M. (1982), 'Inter-industry technology flows and productivity growth', *Review of Economics and Statistics*, 64 (4), 627-34.

Scherer, F.M. (1983), 'The propensity to patent', *International Journal of Industrial Organization*, 1 (1), 107-28.

Schmookler, J. (1966), *Invention and economic growth*, Cambridge MA, Harvard University Press.

Schoar, A. (2002), 'Effects of Corporate Diversification on Productivity', *Journal of Finance*, 57 (6), 2379-403.

Schor, J.B. (1991), *The overworked American*, New York, BasicBooks.

Schumpeter, J.A. (1942), *Capitalism, Socialism, and Democracy*, published in Great Britain in 1943, second impression 1944, London, George Allen & Unwin.

Scitovsky, T. (1954), 'Two concepts of external economies', *Journal of Political Economy*, 62 (2), 143-51.

Scitovsky, T. (1976), *The joyless economy*, New York, Oxford University Press.

Searle, A.D. (1945), 'Productivity Changes in Selected Wartime Shipbuilding Programs', *Monthly Labor Review*, 61 (December), 1132-47.

Segerstrom, P. (1998), 'Endogenous Growth Without Scale Effects', *American Economic Review*, 88 (5), 1290-310.

Servaes, H. (1996), 'The Value of Diversification During the Conglomerate Merger Wave', *Journal of Finance*, 51 (4), 1201-25.

Shapiro, M.D., and D.W. Wilcox (1996), 'Mismeasurement in the Consumer Price Index: An Evaluation', In: B. Bernanke and J. Rotemberg (eds.), *NBER Macroeconomics Annual 1996*, Cambridge MA, The MIT Press, 93-142.

Shleifer, A., and R.W. Vishny (1989), 'Management entrenchment: the case of manager-specific investments', *Journal of Financial Economics*, 25 (1), 123-39.

Smith, A. (1979), *The wealth of nations*, reprinted with revisions (1776), Penguin Classics 1986, London, Penguin Books.

Smulders, S., and T. van de Klundert (1995), 'Imperfect competition, concentration and growth with firm-specific R & D', *European Economic Review*, 39, 139-60.

Soete, L.L.G. (1979), 'Firm size and inventive activity: The Evidence Reconsidered', *European Economic Review*, 12 (4), 319-40.

Sokoloff, K.L. (1988), 'Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790-1846', *Journal of Economic History*, 48 (4), 813-50.

Solow, R.M. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, 70 (February), 65-94.

Solow, R.M. (1994), 'Perspectives on Growth Theory', *Journal of Economic Perspectives*, 8 (1), 45-54.

Steedman, I. (2001), *On 'Measuring' Knowledge in New (Endogenous) Growth Theory*, Paper presented at the Growth Theory Conference, Pisa, October 5-7th.

Stoneman, P. (2002), *The Economics of Technological Diffusion*, Oxford, Blackwell Publishers.

Teece, D.J. (1980), 'Economies of scope and the scope of the enterprise', *Journal of Economic Behavior and Organization*, 1 (3), 223-47.

Teece, D.J. (1982), 'Towards an economic theory of the multiproduct firm', *Journal of Economic Behavior and Organization*, 3 (1), 39-63.

Thompson, P. (2001a), 'How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study', *Journal of Political Economy*, 109 (1), 103-37.

Thompson, P. (2001b), 'The Microeconomics of an R&D-Based Model of Endogenous Growth', *Journal of Economic Growth*, 6 (4), 263-83.

Tobin, J. (1969), 'A General Equilibrium Approach To Monetary Theory', *Journal of Money, Credit and Banking*, 1 (1), 15-29.

Triplett, J. (1987), 'hedonic functions and hedonic indexes', In: J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Vol. 2, London, MacMillan, 630-4.

Van de Stadt, H., A. Kapteyn, and S. van de Geer (1985), 'The relativity of utility: evidence from panel data', *Review of Economics and Statistics*, 67 (2), 179-87.

Veblen, T. (1899), *The theory of the leisure class*, Mentor edition 1953, New York, The New American Library.

Vernon, R. (1966), 'International investment and international trade in the product cycle', *Quarterly Journal of Economics*, 80 (May), 190-207.

Verspagen, B. (1996), 'Technology indicators and economic growth in the European area: some empirical evidence', In: B. van Ark and N. Crafts (eds.), *Quantitative aspects of post-war European economic growth*, Cambridge, Cambridge University Press, 215-43.

Verspagen, B. (1997), 'Estimating International Technology Spillovers Using Technology Flow Matrices', *Weltwirtschaftliches Archiv*, 133 (2), 226-48.

Walsh, V. (1984), 'Invention and innovation in the chemical industry: demand-pull or discovery-push?', *Research Policy*, 13, 211-34.

Wells, L.T. Jr. (1972a), 'International Trade: The Product Life Cycle Approach', In: L.T. Wells Jr. (ed.), *The Product Life Cycle and International Trade*, Boston, Harvard Business School Press, 3-33.

Wells, L.T. Jr., ed. (1972b), *The Product Life Cycle and International Trade*, Boston, Harvard Business School Press.

Williamson, O.E. (1967), 'Hierarchical control and optimum firm size', *Journal of Political Economy*, 75 (2), 123-38.

Young, A. (1998), 'Growth without Scale Effects', *Journal of Political Economy*, 106 (1), 41-63.

Zolotas, X. (1981), *Economic growth and declining social welfare*, New York, New York University Press.

Zweimüller, J. (2000), 'Schumpeterian Entrepreneurs Meet Engel's Law: The Impact of Inequality on Innovation-Driven Growth', *Journal of Economic Growth*, 5 (2), 185-206.