

**Externalities versus Technological Complementarities:
a model of GPT-driven, sustained growth**

by

Kenneth I. Carlaw

Lecturer, University of Canterbury, New Zealand

email: k.carlaw@econ.canterbury.ac.nz

and Richard G. Lipsey

Emeritus Professor of Economics at Simon Fraser University

email: rlipsey@sfu.ca,

home page: <http://www.sfu.ca/~rlipsey>

28 May 2001

**A Paper to be Presented To The
Conference in Honour of the 20th Anniversary of Nelson and Winter's Book
An Evolutionary Theory Of Economic Change
Aalborg Denmark, 12-15 June 2001**

ABSTRACT

Nelson and Winter (1982) emphasised the behaviour of firms as agents of the technological change that underlies long term economic change and growth. In order to establish some key stylised facts about technical change, we concentrate in Part 1 on big technological shocks known as general purpose technologies (GPTs). We argue that *externalities* as conventionally defined miss many of the spillovers that are both causes and consequences of the technological changes that underlie much of economic growth. We then introduce a much wider concept called *technological complementarities*. New GPTs (and other major technological changes) rejuvenate the growth process by creating technological complementarities, most of which are not measured by externalities.

In Part 2, we consider the problem of how to move from the micro behaviour that Nelson and Winter and the present authors have analysed to the kind of macro observations that growth economists seek to explain. We argue that the move cannot be made by formal aggregation techniques whenever the specifications of the micro relations are at all realistic. So the type of appreciative theorising championed by Richard Nelson must be used instead.

To this end, we present in Part 3 a non-standard aggregate growth model that is consistent with our micro observations. It has endogenous, incremental technological changes in applied R&D, which are subject to diminishing returns, interrupted by the occasional introduction of new GPTs. The latter are developed by endogenous basic R&D, which has a payoff that is subject to (Knightian) uncertainty with respect to its magnitude and timing. Although it is technically much simpler than those in the volume on GPTs edited by Helpman, we argue that it captures many of the complex inter-relationships involved in the growth process that were suggested by Nelson and Winter but are ignored in most standard macro growth models. The exercise reveals as a formal proposition that growth based on technological change can be sustained with none of the standard externalities present provided technological complementarities operate.

TABLE OF CONTENTS

I. EXTERNALITIES, AND TECHNOLOGICAL COMPLEMENTARITIES.....	1
General Purpose Technologies	1
Information and communication technologies (ICT)	2
Materials	2
Power Delivery Systems.....	2
Transportation.....	3
Organisational technologies	3
Technological inter-relatedness	3
Externalities	4
Static externalities.....	4
Dynamic externalities	5
Measurement of dynamic externalities.....	6
Technological complementarities.....	7
1. The initiating agent captures the rents associated with downstream applications ..	7
2. No surplus value is created.....	8
3. Decisions that change the long-term trajectory of technological development.....	8
4. Discrete actions.....	8
5. Innovations now in the public domain	9
6. Very long lags.....	9
Definition.....	10
The relation between complementarities and externalities.....	10
Measurement of technological complementarities	11
Examples of technological complementarities	11
Technological complementarities not externalities drive long term growth	15
Measurement of technological externalities	16
Measurement of technological complementarities	17
II AGGREGATION	19
Aggregation and market structures.....	19

Aggregation and end-state competition	20
Appreciative aggregation.....	20
III. A non-equilibrium model of sustained growth.....	21
A Two Sector-Model	22
Simulation of the two-sector model.....	23
A Three-Sector Model	24
Simulation in a three sector model	28
Sustainable growth with decreasing returns to scale	29
IV. CONCLUSION: SUSTAINED GROWTH AND ALL THAT	31

Externalities versus Technological Complementarities: a model of GPT-driven, sustained growth

In their seminal work, *An Evolutionary Theory of Economic Change* (1982), Nelson and Winter stress the micro economic behaviour that lies behind our observations of economic growth at the macro level. In Part I of this paper, we look in detail at some of the micro behaviour with respect to technological linkages that were not considered by these authors. In Part II, we raise the question not addressed by them of how one can go from the sort of micro observations that seem to characterise technologies and technological change in the real world to theories of the sort of macro growth behaviour that growth economists typically seek to explain. In Part III, we build a macro model that captures some of the micro technological relations that we discuss in Part I.

I. EXTERNALITIES, AND TECHNOLOGICAL COMPLEMENTARITIES

The economy of any developed country is characterised by a complex set of interrelationships, which spread the effects of any major technological change far beyond those agents that initiate the change. (We use the term “spillover” to cover all such interrelationships and later we define precise measures that cover these.) In this part of our paper, we study these spillovers. We first discuss the concepts of general purpose technologies, externalities and technological complementarities. Then we compare the last two concepts. Next, we outline several case studies of technological complementarities that show the large amount of spillovers that are missed by measures of externalities. This leads us to several important conclusions. (1) The spillover effects that new technologies have on firms, industries, and sectors beyond those in which they originated are nowhere near fully covered by externalities as conventionally defined and measured; they are covered rather by the much wider concept of technological complementarities. (2) These spillover effects are not, therefore, correctly measured by the estimates of externalities that are commonly found in the literature. (3) Technological complementarities are the vehicle for sustaining the growth process—growth that will at various historical periods be faster or slower depending on the nature of the complementarities associated with successive major technologies. (4) The effects of technological change must be measured by a counterfactual experiment and are not, therefore, correctly measured by total factor productivity.

To develop our analysis, we need to introduce the related concepts of *general purpose technologies*, *technological externalities*, and *technological complementarities*.

General Purpose Technologies

The economic changes that result from technological changes in any growing economy range from the benefit of small incremental improvements in existing technologies through to the big changes in economic structure and performance driven by massive

technological shocks associated with what are called general purpose technologies (GPTs). GPTs share some important common characteristics: they begin as fairly crude technologies with a limited number of uses; as they diffuse throughout the economy, they evolve into much more complex technologies with dramatic increases in their efficiency, in the range of their use, in the range of economic outputs that they help to produce, and in the range of new technologies that they enable. As mature technologies, they are widely used for a number of different purposes, and they have many *technological spillovers* in the sense of co-operating with, and sometimes requiring amendments to, many other technologies, as well as creating myriad possibilities for the invention of new technologies.¹

Lipsey Bekar and Carlaw have studied the enormous impacts on economic, social and political relations caused by some of the most important GPTs from history, identifying about 20 since the Neolithic agricultural revolution.² We have space here only to at hint of their importance.

Information and communication technologies (ICT)

Writing invented in Sumer around 3500-3200 BC, increased by orders of magnitude the degree of complexity with which societies could be organised. *Printing* with movable type enabled many of the dramatic changes that occurred in Europe between 1500 and 1700. The current *Information and Communication (ICT) Revolution*, driven by several related technologies grouped around the electronic computer, is changing product design, production, marketing, finance, the organisation of firms, and creating a wide range of new products incorporating hard coded chips, computers and/or software. The *Internet*, which was enabled by the computer (which was in turn enabled by electricity), is a GPT in its own right.

Materials

Bronze, which gave its name to an age lasting for about 15 centuries starting about 2800 BC, revolutionised warfare allowing the emergence of multi-city empires and colonial wars; it led to the emergence of market societies, which superseded the command societies that had grown up around writing; it led to the replacement of the priests by kings as the predominant political leaders. *Iron* produced a revolution in civilian and military technologies that was accompanied by the destruction of every major city from Crete to the Egyptian border; events that initiated a dark age that lasted several centuries. The current revolution in materials technology, allows *made-to-order materials* to be invented specifically for use in newly developed products and processes.

Power Delivery Systems

The Water Wheel was invented in Roman times but its diffusion throughout early Medieval Europe led to the early mechanisation of significant amounts of European manufacturing. *Steam power* facilitated the creation of new products, new production

techniques and, eventually, many new industries, slowly replacing sails at sea and rapidly replacing long-distance horse transport on land. *Electricity* fundamentally changed the world, taking it from mechanical to electronic sources of power and allowing the invention of a vast range of new products and processes—an evolution that continues to this day. The *internal combustion engine* enabled the *automobile* and the *lighter-than-aircraft*—two of the 20th century’s most important transportation GPTs.

Transportation

The *three masted sailing ship* allowed Europeans to sally forth to trade with the whole world, to conquer much of it, and to destroy many of its civilisations. *Railways* were the first mobile, inanimate power delivery system for land uses and, in conjunction with the iron steam ship and the telegraph, they created a single market across the whole United States, opened up vast parts of the world to settlement, and allowed the production of foodstuffs for distant markets. The *motor vehicle* has profoundly affected the 20th century economy in both production and consumption. The *commercial aircraft*, like most GPTs began with limited uses but once the jet was firmly established, the era of mass travel, nation-wide sports leagues, and a host of other communication links, encouraged the globalisation of markets.

Organisational technologies

The *factory system*, established by the First Industrial Revolution, was a response to particular innovations in capital goods that required different organisations at all levels, including finance, and different amounts of human capital leading to the decline of villages and the growth of industrial cities. *Mass production* largely eliminated craft production and spread from the auto industry to cover most American manufacturing, making it, by the mid 20th century, the dominant method of organising manufacturing in North America, and to a lesser extent in Europe. *Flexible manufacturing*, lean production, or Toyotaism, as it is variously called, was driven by the need of Japanese automobile makers to produce competitively at a scale much less than was possible for US firms. It largely eliminated the system of mass production as its principles spread from the shop floor to every phase of the firm’s operations, from design, to marketing, to customer follow up.

Technological inter-relatedness

As GPTs diffuse through the economy, they create research programs for entrepreneurs to apply the GPT’s principles to create new processes and new products. This creates a complex web of relations as developments in one firm, industry and sector influence developments in other firms, industries and sectors. Think, for example, of all the myriad ways that innovators have found to use electronic chips; how these ways have multiplied as the power and reliability of chips have increased; and how some of these ways have in turn enabled other developments; and so on in a complex concatenation of derivative innovations. To capture these and other interrelationships,

economists typically use the concept of an externality. We first discuss this concept and then go on to argue that the interrelationships are better captured by the much more embracing concept of technological complementarities.

Externalities

As it is normally understood, an externality is an unpaid-for effect conferred by the actions of one set of agents on another set of agents who are not involved in the activities of the first set. We refer to these agents as the “initiating agents” and the “receiving agents” respectively. Externalities typically prevent the fulfilment of the optimality conditions since there will be too much of activities that produce negative externalities and too little of the activities that produce positive externalities. We confine ourselves to the analysis of positive externalities.

Static externalities

The standard text-book treatment of externalities takes place in a GE framework in which tastes and technology are fixed and there is no irreversible “arrow of time”. A positive production externality is usually analysed as an unpaid-for input into the receiving agents’ production function that is created by the initiating agents’ activities. The typical formulation is illustrated by the following case:³

$$(I.1) \quad A = ax_a^\alpha y_a^\beta \quad \alpha + \beta = 1$$

$$(I.2) \quad B = bAx_b^\gamma y_b^\delta \quad \gamma + \delta = 1$$

where A and B are the outputs of two products, x_a, y_a are the input choices for agents producing A and x_b, y_b are the input choices for agents producing B . The action of A agents to maximise the objective function I.1 has a positive impact on the B objective function I.2.

Because we wish to deal with technological change that takes place in irreversible time, we need to state the definition of externalities a little more explicitly than is usual:

Externalities are unpaid-for effects conferred by the *continuing* and *potentially variable* actions of the initiating agents on receiving agents who are not involved in the initiating agent’s activity, and for which the receiving agents would be willing to pay.

Within this broad class we distinguish two subclasses, which we call static externalities and dynamic externalities.

Static externalities are externalities as defined above that occur when the actions of the initiating agents and the effects on the receiving agents are those that can occur in an Arrow-Debreu type GE model with constant tastes and technology.

Note the following characteristics of the analysis of externalities in the typical text-book treatment.

1. The action takes place in a standard GE formulation in which tastes and technology are constant (i.e., they are static externalities).
2. The initiating agents' actions are continuous and potentially variable.
3. The receiving agents would be willing to pay something to alter the initiating agents' actions and hence, if they cannot do so, optimality is not achieved.
4. The receiving agents accept the externality passively; it provides them with an input or something similar that does not require any further action on their part to make it useful to them.
5. The variables are undated so that the impact on the receiving agents is implicitly assumed to be contemporaneous.

Dynamic externalities

Following Arrow (1962), externality analysis has been extended to apply to a world in which technology is changing. Because knowledge is non-rivalrous, a technological innovation typically confers benefits on many agents beyond the firm, industry and even sector in which it was originally made. This gives rise to what we call dynamic externalities

Dynamic externalities occur when the continuing action of initiating agents generates technological changes that affect the actual or potential technologies of the receiving agents; the value of existing technologies owned by the receiving agents may be increased⁴ and/or opportunities may be created for the receiving agents to make further technological changes. Since externalities are normally limited to actions that would upset optimality conditions, for an action involving technological change to give rise to a dynamic externality, it must be one that a social planner would alter to achieve a better outcome and/or the affected agents would be willing pay to have altered.

Listed below are five characteristics of the analysis of dynamic externalities that parallel those listed above for static externalities, followed by a comparison between the two lists.⁵

1. Technology is allowed to vary.
2. The initiating agents' actions are continuous and potentially variable.
3. The receiving agent would be willing to pay something to alter the initiating agent's actions and hence, if they cannot do so, optimality is not achieved.
4. The receiving agents do not necessarily accept the externality passively; it may provide them with an opportunity to engage in potentially profitable R&D that was not present formerly, the benefit of the externality coming only after the new R&D produces a profitable innovation.

5. The variables may be dated to allow for a lag between the actions of the initiating agent and the effect on the receiving agent.

Characteristic 1 is altered by the definition of a dynamic externality.

Characteristic 2 is unchanged. The initiating agents' actions must be continuing and alterable. If they are complete, they are by-gones. Even if inventions resulting from past actions are used in current production, they have no potential to upset optimality conditions since they are fixed and unalterable.⁶ If they are continuing but cannot be altered, they are a background against which maximisation can take place, just as are all the legacies of the past.

Characteristic 3 is unchanged since it is an essential characteristic of an externality.

Characteristic 4 is altered since affected agents may not be presented with something it can directly use, such as a free input; instead the externality may take the form of the opportunity to do new R&D. For example, the invention of the dynamo and the development of an electricity generation and distribution infrastructures presented the durable consumer's goods industry the opportunities to develop the range of electrical machines that revolutionised household management in the period between the First and Second World Wars—refrigerators, electric stoves, vacuum cleaners, washing machines, deep freezers, electric mixers, dish washers etc.)

Characteristic 5 is altered since, when we are considering the impact of technological innovations on agents other than those who initiate them, we must allow for the fact that impacts are often spread over a long time period—as when an innovation in industry A becomes sufficiently well established to provide an opportunity for industry B to do profitable R&D to use some aspect of A's developments. For example, it was some time after the invention of the high pressure steam engine that railroad building began in earnest.⁷ steam ships began dynamo that the R&D to develop the household electrical machines referred to above began.

Measurement of dynamic externalities

Since inventions and innovations are discrete events, both the theoretical and empirical analysis of the dynamic externalities that are associated with them are usually carried out using the R&D expenditure that helps to create them, rather than the inventions themselves.⁸ The usual assumption is that the relevant stock of accumulated (and depreciated) R&D expenditure is an input into a firm's production function. The externality then arises because the R&D accumulated by initiating firms enters into the production functions of receiving firms. Sometimes the externality is assumed to be totally general so that the nation's aggregate stock R&D is entered into each firm's production function.

In order to allow us to make contrasts between externalities and complementarities, we need to outline the two methods have been used in attempts to measure dynamic externalities.⁹ One method is to look for a correlation between the total factor productivity (TFP) or labour productivity of a firm, an industry or a sector on the one hand, and some measure of R&D investment "outside" that firm, industry or sector, on the other hand. Here the R&D is often taken to be the current flow of expenditure on

R&D weighted by output (R&D intensity) not the accumulated stock of R&D. The parametric estimate is the gross rate of return to R&D, which in theory is output divided by the input of R&D capital weighted by the estimated elasticity of output with respect to R&D capital (i.e., the marginal product of R&D). A rate of return that rises as the level of aggregation rises is taken as evidence of positive spillovers. This is a test for contemporaneous free gifts in the form of output not matched by paid-for inputs. (Mairesse and Sessenou (1991) provide a survey of this literature.)¹⁰

A second method is to calculate the elasticity of output with respect to R&D at various levels of aggregation. Here the R&D input to production is usually measured as the stock of R&D, which requires treating R&D as a normal input into current output. There is a stock of knowledge capital that is internal to the firm and an outside stock of knowledge capital that benefits the firm thus providing an externality. In the standard Cobb-Douglas formulation, this elasticity is the estimated exponential parameter on the R&D input. At the firm level the elasticity with respect to the internal stock of knowledge capital will be smaller than the elasticity with respect to the entire stock that actually gets used by the firm. An increase in the measured elasticity of the knowledge stock as the level of aggregation increases is taken as evidence of a positive R&D spillover.¹¹ (Griliches (1982))

Technological complementarities

In the typical case of a technological spillover, an innovation made by a set of initiating agents, A, creates opportunities for a set of receiving agents, B, to make other innovations that incorporate, or in some way rely on, the initial innovation. A's innovation does not directly shift B's production function. Instead, it creates an opportunity for B to conduct R&D to invent a new technology and invest in installing it. This will eventually alter B's production function, either by altering the relation between inputs and an unchanged product or by altering the nature of the product itself. The histories of the GPTs that we listed earlier are full of examples. These spillovers will sometimes give rise to dynamic externalities as defined above but in the following circumstances they will not. (These circumstances were mostly suggested to us by the study of actual cases of specific technological advances. In presenting them, it seems best to reverse the order of our own development by giving the generalisations first and then illustrating them with specific examples—although the examples are needed to show that these are common circumstances, rather than arcane theoretical curiosities.)

1. The initiating agent captures the rents associated with downstream applications

Dynamic externalities are created only to the extent that the initiating agents in A cannot capture all of the benefits conferred on the receiving agents in B. The initiating agents may capture none of the benefits (all benefits are external to A), or some of the benefits (some of the benefits are external to A), or all of the benefits (none of the benefits are external to A). This latter case would be most likely if the downstream applications of the new innovation were specific and localised. But whatever the extent of the externality,

the spillover is there in the sense that something done in *A* is useful in *B*. whether or not the spillovers give rise to externalities will depend on such things as existing property rights, the market power of the initiating and receiving agents, and the degree to which downstream benefits are anticipated.

2. No surplus value is created

Let the revenue arising from the receiving agent's innovations that were enabled by the initiating agents' original technological advance just cover the R&D costs of developing them and yield a rate of return on new capital investment just equal to what would have been yielded by investing in existing technologies. In this case, there is technical advance but the cost of making it just covers the opportunity cost of the alternative of investing in existing technologies. There are then no rents or "free lunches" to be gained and hence the receiving agents would not pay the initiating agents to do more of their R&D. Thus there are no externalities but there are technological spillovers that provide receiving agents with opportunities that they did not previously have and that are worth taking up. (This is the situation emphasised by Jorgenson and Griliches (1967) when they argue that growth need not be associated with the kind of surplus that is measured by total factor productivity.)

3. Decisions that change the long-term trajectory of technological development

Agents in different countries may make decisions that put their countries on different specific trajectories of technological development. These different decisions may each have been maximising ones, given such things as international differences in natural resource endowments, climate, and available human skills. Importantly, different technological trajectories have different, often unforeseeable, potentials for further technological advances. The realisations of these potentials create spillovers from the static choices of techniques to the dynamic path of technological change. But they are not necessarily externalities because, depending on the rate of time discount, agents or social planners, presented with perfect foresight, might still prefer the path that maximised with respect to current conditions, even though it led to a slower rate of technological advance than would an alternative path (that would be based on a statically non-optimal allocation of resources).¹²

4. Discrete actions

Inventions and innovations are discrete actions as opposed to continuous ones. Although industry *A* may take a long time perfecting an invention before it becomes marketable, its impact on industry *B* occurs all at once when the invention becomes available to it.^{13 14} To show a pure case, assume that industry *A* has no significant impact on industry *B* except for one discrete innovation in *A* that has a major effect on *B*. That innovation comes in the middle of a long trajectory of *A*'s technological advances that used a developing body of general engineering knowledge. Even if all those agents in industry *B*

who benefit from this one innovation could have been brought together before A's innovation was made and given perfect foresight, there is little that they could have done to alter A's behaviour so as to increase their benefit. A did the necessary R&D at the appropriate time in the trajectory of all of its own developments, as well as those engineering advances that it used from outside, and it did not suffer capital rationing. For all of these reasons, more R&D funds would not have altered the nature of its innovation. Nor could they have done much to its timing. The innovation was part of a trajectory of related discoveries in industry A and elsewhere. Possibly, a massive injection of money would have hastened the development by a year or two but no more. The details will vary case by case, but the key points are that because A's innovation is part of a developing set of linked innovations, each building on the others, as well as engineering knowledge being developed elsewhere, there is little scope for B agents to accelerate the innovation from which they gain, even if they knew what was happening in A and had perfect foresight about its effects on them.

5. Innovations now in the public domain

Agents continually enjoy the free use of all previous technological advances that are now in the public domain and still in use. Anyone who uses the wheel, the lever, the water wheel (embedded, for example, in an electricity-generating turbine), sandpaper, steel, screws, computers, and a host of other technologies invented in the past, gains from them. Agents often incorporate one or more of these existing technologies into some newly invented technology that would be impossible without them. For example, there could be few machines if there were no wheels and axles and there could be no personal computers if there were no electricity. The economic durability of many basic technologies gives rise to an enormous list of such unpaid-for technological interactions.¹⁵ Many agents would pay to avoid the alternative of doing without one or more of these public-domain technologies. But these are not externalities since they are not the result of current activity that a social planner would wish to alter. Nor would the set of current beneficiaries, transported back to the time of the innovation and given perfect foresight, want to pay present-day resources to accelerate or otherwise alter the innovative activities before returning to the present—because an alteration in that long past activity would have no appreciable effect on current opportunities. For example, advancing by a century the invention of such important technologies as the wheel and axle, the lever, the screw, the heavy plough, or the spinning jenny would have so little effect on current values that present-day agents would not be willing to pay to reap the present benefits of such an alteration in history.¹⁶

6. Very long lags

Even if the R&D of the initiating agents is continuous and not yet in the public domain, its effects on the receiving agents may be spread over very long periods. Even if recipients would be willing to pay to encourage the original innovating activity, the benefits are so spread out in time that no contemporaneous measure of externalities will

catch them. This gives rise to a new category: things that conceptually are externalities but that can never be measured as such. The extreme case is when one GPT enables another as for example when electricity, enabled computers, which in turn enabled the internet, much of biotechnology and even more of nanotechnology.

Definition

These considerations lead us to define a new concept:

A technological complementarity arises in any situation in which the past or present decisions of the initiating agents with respect to their own technologies affect the value of the receiving agents' existing technologies and/or their opportunities for making further technological changes.

The following is the list of this concept's characteristics that are parallel to those of technological externalities as defined above, and a comparison with technological externalities:

1. Technology is allowed to vary.
2. The initiating agents' actions may be discrete and/or long past and hence non-variable.
3. The receiving agent may be unwilling to pay for the spillover either because the initiating agent appropriates all of the extra value or because there is no surplus value, since the receiving agents' enabled innovating activity just recovers its full opportunity cost. (An "enabled innovating activity" is one that receiving agents could not have undertaken without the innovation of the initiating agents.)
4. The receiving agents do not necessarily accept the externality passively; it may provide them with an opportunity to engage in potentially profitable R&D that was not present formerly
5. The lags may be so long and/or the activity so localised in time that no practical measure of externalities will capture the spillovers.

Characteristics 1, and 4 are the same for technological externalities and complementarities. The other three characteristics are different for reasons made obvious in the previous discussion.

The relation between complementarities and externalities.

Without technological complementarities, the non-rivalrous nature of technological knowledge would be unimportant for growth and there would be few, if any, technological externalities. In other words, if every final good and service were unique, with its own unique production process and its unique set of intermediate goods, new knowledge would be user-specific without spillovers to other goods and production processes. *Therefore, the existence of a technological complementarity is a necessary condition for the existence of a technological externality.*

However, we can have technological complementarities without externalities. The six situations mentioned above provide the reasons. *Thus the existence of an externality is not a necessary condition for the existence of a significant technological externality.*

In summary, let C be the class of all technological complementarities (and C' its negation) and E the class of all externalities (and E' its negation). Then, $C' \cap E$ defines standard externalities; $C \cap E$ defines technological externalities; and $C \cap E'$ defines those technological complementarities that do not give rise to externalities and so are not covered by conventional measures of the external effects of technological change that rely on measuring externalities.

Measurement of technological complementarities

Because technologies interact, co-operate with, and support each other, (i.e., they are complementary), it is in principle impossible to measure the amount that each one separately contributes to the value created by the group as a whole. Consider an agent faced with the choice of doing without the first item on a list of all the existing technologies that he uses freely, then doing without only the second item, and so on down through the list of all the technologies that the agent uses. The sum he would pay to avoid doing without each of these taken one at a time will typically vastly exceed all the current value the agent is able to create using all of them.

Nor can the responses of one technology to a change in another cannot typically be modelled (for measurement or any other purposes) as the consequence of changes in the prices of flows of factor services found in the original production function. This is because all of the action is taking place in the structure of capital.¹⁷ For example, the mass production factory that was introduced after the invention of machine tools that could cut hardened steel (see later discussion) could not have been made possible by a fall in the cost of making parts with the old machine tools that could only cut softer metals, even if that cost fell to zero. The new possibilities depend on how one technology is related to another, not on how a given technology can respond to a change in a price signal.

Examples of technological complementarities

We now put some important observational flesh on the bare bones of our definitions. We wish to show the importance of technological complementarities in actual growth and to illustrate how sustainable growth can be driven by technological innovations that create spillovers many of which are not externalities.

Example 1 American railways: As soon as Watt's patent on the steam engine expired in 1800, there was a burst of R&D to develop the high pressure engine that was needed to allow steam engines to escape from the stationarity imposed by the high weight/power ratio of low pressure engines. The high pressure engine made possible, among other things, the steam ship, first seen on American rivers, and the railway, first seen in Britain. Part of the value of the spillovers that the high pressure engine created for the railways was appropriated by the engine manufacturers (situation 1 above). However, the development of the railways was much more costly than

expected, with high costs of R&D, learning by doing and using, and high costs of fixed capital. Failures were frequent. So although individual fortunes were made, the overall investment in railways gave only modest returns. This is situation (2) above.

American railways in their turn had many spillovers, including the creation of a single US market that made scale economies available to US manufacturers that were unavailable in Britain, and the development of complex forms of management control that later proved useful in many other industries.¹⁸ The development of a system of railways was extremely important to US (and British growth). But this was because it created new opportunities for investments in railways that earned little more than opportunity cost but which conferred further spillovers onto much of the rest of the economy. These spillovers, such as the creation of a single unified market over the whole of the US, are examples of case 5. The lag between the building of the railways and the exploitation of scale economies by numerous affected US industries was too long and diffuse to show up as a measurable externality received from the railroads by other industries. (Furthermore, much of the benefit was not associated with R&D in the railroad industry but with the fixed investment in building them and learning by doing and learning by using, which although a kind of R&D, is not included in conventional measurements of spillovers that associate explicit R&D in one place with total factor productivity elsewhere. This raises the distinction between formal (measured) R&D and informal R&D, which to the extent it gets measured is measured as other inputs. This is another empirical distinction between complementarities and externalities.)

Example 2 The American System of Manufactures¹⁹: The US natural resource base (along with other key forces on the demand side) encouraged the development of technologies that the 19th century Europeans called “The American System of Manufactures”. These resource intensive and labour saving techniques were based on standardised goods produced by specialised machines. While this system developed in the US through the second half of the 19th century, Europe’s resource endowments encouraged craft techniques that were labour intensive and resource saving. It turned out, *in ways that were not, and probably could not, have been foreseen*, that the US system was capable of much more technological improvement than the European craft-based system. So, in an excellent example of the genuine uncertainty that accompanies technological change, the US choice of techniques, based on static advantages, accidentally conferred on the US the dynamic opportunity for much more technological change than did the European choice. The technological complementarity here is that the choice of one basic type of technology, the American system in his case, can greatly enlarge the space for radical and incremental innovation compared to the space made available to those choosing alternative methods of production.²⁰

This is a clear example of situation 3 above. This choice of path, while creating higher growth and higher incomes in the US compared with the UK for over 100 years, need not be reflected in measured externalities. Neither of the conventional measurements of looking either for a correlation between one firm’s output and another’s R&D or for a higher return to R&D at the industry level than at the firm level, will capture the fact that all US manufacturing industries had more, and more profitable, opportunities for investing in new innovations than did British

industries, and that these occurred over the course of more than 100 years after the basic choice of technique had been made. Indeed, there is no externality here at all since no American would have paid to alter the static maximising choices that agents in the United States made and that gave rise to the American system of manufacturing.²¹

Example 3 Automobiles²²: Early machine tools could only cut soft metals. These were then hardened after manufacture, a process that warped the parts that then had to be filed individually to make them fit. Early in the 20th century, machine tools were invented that could cut hardened steel. Henry Ford tried to get his suppliers to work to the high degrees of tolerance made possible by the new machine tools and, when they were unable or unwilling to do so, he took parts manufacturing in house. Having achieved production of a standardised product based on interchangeable parts, it was a small step to introduce the assembly line, which created genuine mass production. (The factory assembly line could not have been introduced in a steam plant where each machine had to be arranged in the relation to its power consumption, while it was made possible by the unit-drive, electrically driven machine tools, which could be arranged in the order that they were needed in the production process.) Ford's methods spread to virtually all other assembly operations in US manufacturing.

This is an extreme example of situation 4 above. One small invention in machine tools had ramifications that transformed first the automobile industry and then virtually all manufacturing.²³ Whether or not there were parallel massive externalities depends on where the value was created and in whose hands it ended up. Certainly, the machine tool manufacturers who innovated these new techniques received only a small portion of the new value that their invention created across all manufacturing. Much of the value was appropriated by Henry Ford and his workers. But his high returns should be seen as a return to having the foresight to see the potential of the new inventions and the ability to realise that potential when current parts manufacturers were unable to do so. This is a good example of situation 4. Since the enabling R&D was done in the machine tool industry around 1900, the payoffs that extended for the next 50 years would not have shown up in any measure of externalities based on contemporaneous costs and benefits. Strictly, there is no externality here because the R&D in the machine tool industry followed its own trajectory of successive innovations that built on each other. Even if all those who benefited from this one innovation could have been brought together in 1900 and given perfect foresight, it is not clear what they could have done to alter the behaviour of the machine tool industry. Possibly a massive injection of money would have hastened the development by a year or two, possibly not.

Example 4 Electricity²⁴: Electricity is a GPT that made possible a vast range of products, processes and organisations that permeate today's entire production system and that is still giving rise to myriad innovations that could never have occurred in a world of strictly mechanical technologies. In such a mechanical world, the electrification of factories would never have occurred. Telephone, telegraph, radios, TV, faxes, electric lighting, email, Internet, satellite signals, and all of the other electronic technologies that go to create the modern ICT revolution would not have existed. Vacuum tubes, silicon

chips and electronic computers of all sorts would have been unthought of, along with most of the household gadgets that revolutionised household management in the first half of the 20th century. Clearly, electricity created an enormous set of technological complementarities in the sense that it created valuable opportunities for new product and process innovations and new investments in virtually all of the nation's industries, old and new.

Furthermore, electricity enabled the development of subsequent GPTs. Electronic computers are on a par with electricity in the widespread revolution that they have enabled in almost all economic relations. Similarly, many of the major developments in biotechnology, which is well on its way to becoming the next GPT to profoundly affect the economic and social structure, and most of nanotechnology, which may revolutionise all production by the middle of the 21st century, would both have been impossible without the computer, which in turn was enabled by electricity.

This is an example of situations 5 (the dynamo has long been in the public domain although its spillovers continue to drive much technological change), situation 6, the lagged effects stretching in this case well into their second century and showing no signs of abating, situation 2, many of the subsequent innovations enabled by electricity returned only their opportunity cost to investors, yet they were a part in an evolving sequence of related technological changes. Indeed, the vast set of interrelated technologies that were enabled by electricity transformed the society and fed the growth process for decades and this happened independently of who appropriated that value (it was often the innovating industries themselves) and how much return there was in each investment in an enabled innovation over and above the rate that could have been obtained by investing in existing technologies. Even if all the people who benefited from the invention of the practical dynamo could have been brought together and given perfect foresight in say 1850, it is not clear how much they could have done to accelerate the invention of the dynamo and the subsequent evolution of electricity. Some, but probably not too much, after all the dynamo was the end product of a trajectory of research into magnetism and electricity that stretched over centuries.²⁵

Example 5 Machine tools²⁶: Rosenberg's work on the US machine tool industry analyses some important examples. Most of the general purpose machine tools developed in the US in the 19th century, such as the turret lathe and the universal grinder, were developed to solve problems in particular industries, firearms in these two cases. Later, many of the key innovations in machine tools were used to solve problems in sewing machine and bicycle production. The machines developed for these purposes created enough value in the industries to which they were originally developed to pay for their development costs—as shown by the fact that the firms that developed them remained profitable before they expanded their sales into other industries. The machine tools went on, however, to become highly useful in most metal using industries—the reason being that all cutting and shaping of metal requires relatively few basic operations, most of which are common to all metal users. This is a process that Rosenberg calls technological convergence, by which he means that industries producing unrelated final products

evolve to use common process technologies. These were important technological complementarities, which probably are reflected more in measurable externalities than those any of our of cases. There are, however, several reasons why some of the complementarities would not have shown up as externalities. First, the machine tool producers who developed the new tools were often able to sell them to users other than the original ones at prices that internalised some of the extra benefit over cost that the use of these tools created (situation 1). Second, the application of the new tools to uses other than those for which they were designed required some developmental costs including R&D (situation 2). Third, these and other similar tools, remained an input into the innovative process over subsequent decades (situation 6) and, even today, the concept and uses of these tools, now in the public domain, is taught in trade schools and some people have them in their garage (situation 5). Nonetheless, this is a case in which there were probably significant measurable externalities as the new machine tools lowered costs in many user industries. If these industries were competitive, their consumers would have reaped the gain in terms of lower prices, but the user industries would have recorded high TFP growth as the new machine tools lowered their costs.

Technological complementarities not externalities drive long term growth

The whole process of economic growth can be seen as being driven by a succession of GPTs sometimes overlapping and sometimes discretely separated by long periods in which only small incremental changes are being made in well established GPTs. In short, the growth process is rejuvenated by the broad-based R&D program that slowly evolves from each new GPT. This process has major discontinuities in the sense that each R&D program could not have evolved incrementally from the one that preceded it. For example, the economy-wide R&D program for applying electricity was very different from that for applying the steam engine.

Thus the technological change that is actually observed at the micro level is neither smooth nor balanced. It does not cause continuous variations in the marginal productivity of capital in each firm and industry as more capital is accumulated and as technological change slowly and continuously alters the relation between inputs and outputs. Instead, major technological advances cause discontinuities in the opportunities for new investment in particular sectors and radically alter the relation between inputs and outputs at the microeconomic level. The reason economists do not observe such discontinuities when they fit a single macro production function is that new GPTs start in isolated sectors of the economy and evolve only slowly to assert their dominance over the old technologies that they challenge and eventually largely replace—a process that can easily take half a century or more. Since at any one time, there typically exist several GPTs, each at different stages of their evolution, the macro production function, which is some kind of unspecified aggregation of the functions for individual GPTs, does not show the discontinuities that characterise individual firms, industries and sometimes whole sectors. The level at which discontinuities would be observed depends on the nature of the change under consideration. As with most aspects of technological change, there are few

generalisations here that are universally applicable. When some major new technology spreads slowly from firm to firm, discontinuities will only be observed at the firm level. Even there, radically new technologies often perform little better than those they replace for quite long gestation periods during which learning by doing and learning by using must occur. In such cases, each firm's productivity will change continuously. What this discussion illustrates once again is that observing realised changes in productivity at the aggregate level is a poor way to gage the extent to which technological changes are actually occurring.

To stress the importance of the technological complementarities introduced by any major new technology, especially by a new pervasive GPT, consider two situations.

In the first situation capital is accumulated but there is no technological advance. Diminishing returns to capital would soon set in for all the reasons mentioned in the earlier discussion. In the second situation, there is a succession of GPTs, such as automated textile machinery, the steam engine, the American System of Manufactures, the electric dynamo, the internal combustion engine, the automobile, and the computer. Each of these innovations widens the space in which the search for further innovations could fruitfully take place. One could not, for example, invent the radio and television without electricity, nor advanced robots and the Internet without the computer. Even if the development cost of each of these new technologies that were enabled by a new GPT were just covered by sales revenues, the marginal product of capital would be higher than it would have been under conditions of static technology. *So whether or not there are externalities in the form of technology transfers for which the recipients would have paid more for than they actually did pay, and whether or not there is a discrepancy between private and social rates of return, the technological complementarities that arise from radically new technologies have been a major (we would say the major) source of growth over, at least, the last three centuries.*

Measurement of technological externalities

Earlier we mentioned the two methods that have been used to measure the externalities that are commonly thought to be the manifestation of technological spillovers. Many measurements have been made indicating varied evidence of R&D spillovers at the aggregate level. Griliches 1982 provides an overview of measures indicating that while the majority of studies find some evidence for R&D spillovers, some do not and that there are a number of methodological and empirical problems associated with such measures. As Mairesse and Sassenou (1991) point out, firm level estimates of the rate of return to R&D are as likely to reflect social return as private return. These problems do not surprise us. Our own analysis strongly suggests that both of these measures are based on incomplete conceptualisations of how the effects of R&D, and other sources of technical change, spread beyond those who are actually responsible for it.

We have shown that, in our conceptualisation, what major GPTs such as electricity—and occasionally by a very specific innovation, such as the machine tools that could cut

hardened steel— did was to create opportunities for profitable investments in a large set of new product, process, and organisational technologies. Many of these come into existence long after the original R&D to develop the basic technology has been paid for and forgotten (e.g., the Voltaic cell and the dynamo in the case of electricity). Even if there are no free gifts; no externalities; no super-normal returns on investment in these new technologies,²⁷ there was, and is, enormous economic gain because the opportunities for these investments *would not have existed* without the GPT. Put another way, *GPTs, such as electricity and the computer, expand the space of possible inventions and innovations, creating myriad new opportunities for profitable capital investments, which in turn create other new opportunities, and so on in a chain reaction that stretches over decades, even centuries.*

We have also observed that direct attribution of value to any specific components of technology is impossible. As already observed, because technologies interact with each other in a complex web of complementarities, there is no unique meaningful measure of the value of each. Specifically, the conceptual procedure of eliminating them one at a time and summing the loss in value in each case does not work (as it will not with any set of complimentary inputs)—and typically will produce a value in excess of total GDP.

Measurement of technological complementarities

We have seen that many of the technological complementarities that lie behind sustained growth are not captured either conceptually or empirically by conventional measures of externalities. New technologies allow things to be made or done that could not have been made or done before. In such cases, there is often no satisfactory way of directly measuring the gains resulting from the replacement of the old by the new technologies. Could we measure them by equivalent price changes? For example, we might think of comparing the steam engine and the electric motor with some sort of hedonic index that relied on horsepower or BTUs produced by each motor for equivalent amounts of inputs. But, as we have argued earlier, the economic gain that came when the electric motor replaced the steam engine did not stem mainly from any ability to cheapen the cost of energy. It came, rather, in its ability to permit a reorganisation of production in ways that were technically impossible with steam power. (This illustrates the point made earlier that it is impossible to model the introduction of a new technology that has large technological complementarities as if it were a fall in the price of the old technology.)

This discussion suggests that to measure the benefits of technological change, we should not look to some assumed discrepancy between current private and social rates of return but to a discrepancy between actual returns and what returns would have been had the change not occurred.

Consider the extreme case of an embodied technological change that produces no free lunches but just covers its opportunity costs. Technological advance could not be less productive than in that. (If it were less productive, investment in more of the old technology would be more profitable than investment in the new.) In this lower-limit case, where returns to the new technology just cover the full opportunity cost of

developing and producing it, the rate of return on investing in the new technology will just equal the rate of return on investing in the old. The social advantage of the new technology over the old is then in the *future path of returns*. With the opportunities created by the new technology for further technological innovations that stretch over future decades, the actual rate of return may hold constant instead of falling as it would if technology had remained static.

An illustrative example is presented in Figure 1. (All of the figures are presented at the end of the paper.) Current technology gives a marginal product curve for capital of MP_k' and, with a stock of capital K_1 , the marginal rate of return on investment in old technology is 10%. If technology were held constant, diminishing returns to the accumulation of capital would drive that rate of return down along the curve MP_k' . Further assume that investment in bringing a new technology to market also earns a return of just 10%—there are no free lunches. Investment in the new technology takes place and, as well as yielding 10% for its owners, it shifts that marginal product curve for capital by creating the opportunity for a new innovation in period $t+1$ that also yields 10% as shown by the curve MP_k'' . If technology were held constant from then on, the return to investment would fall along the curve MP_k'' . But the new technology introduced in period $t+1$ creates the opportunity for investment in yet another new technology in period $t+2$ that also yields 10%, which is the same as the yield on period $t+1$'s technology, and is shown by the curve MP_k''' . And so it goes period by period, with each new vintage of technology only earning its opportunity cost, but also creating the opportunity for yet a further new vintage of technology that does the same. The trajectory of the economy's actual investment and returns follows the arrowed path T. At any one time investment in shifting the MP_k curve outwards just equals the return to investing in the existing technology. (Of course the MP_k is meant to be much flatter and the shifts much smaller than they are shown for visibility.)

How then should we calculate the gains from technological change? First, if we compare the returns to investment in R&D with the return for investment in existing technologies at the margin period by period, we will show no gain from technological change since both values will hold constant at 10%. Second, if we compare the returns to R&D period by period with what the returns would have been if technology were unchanged from period t onward, we will find the gains rising period by period. For any base period, say t , the gains are given by the area between the arrow T and the MP_k curve for the technology of the base period. This is the cumulated difference in the return to capital in a technologically dynamic world and what it would have been in a world of static technology.

The important message is that there needs to be no observable impact of the new technology on rates of return; instead the impact is between what actually happens to returns over some future time period and what would have happened in the absence of the technology. The need to make this counterfactual observation makes it difficult to observe the effects of major innovations on rates of return. Nonetheless, the benefit grows over time as the gap grows between what the rate would have been as it fell

continuously under the impact of capital accumulation and constant technology and what it actually is—possibly holding constant or even rising as more and more possibilities inherent in the new GPTs are realised. Think, for example, what the rate of return on investment would have been if we were still investing in more and more capital goods of Victorian design rather than the capital we now create.²⁸

In section III we introduce a model that mirrors this view of the impact of technological change with neither externalities nor free lunches. In the meantime, we give brief consideration to the problem of aggregation.

II AGGREGATION

Both neoclassical and endogenous growth models employ an aggregate production function. Conceptually, it can be aggregated from two theoretical microeconomic worlds, one where all firms are perfectly competitive, the other where they are monopolistically competitive as given by the Dixit Stiglitz model. But the function cannot be formally aggregated from the mixture of monopoly, oligopoly, monopolistic and perfect competition that characterises industrial structures of the real world. Any economy-wide production function is thus a theoretical *assumption* with no direct empirical counterpart in actual micro data. We elaborate on this point below.

Aggregation and market structures

First, consider perfect competition. The world that creates economic growth and technological change is not one of perfectly competitive equilibrium. So being able to aggregate from micro-behaviour to a macro production function in the imaginary world of perfect competition is not particularly relevant here.

Next consider monopolistic competition. The Dixit-Stiglitz model of monopolistic competition is a model that fits few if any markets that we see in the world. Indeed, it has all the features that had come to discredit Chamberlin's model in the eyes of I.O. economists before the revival of the Dixit Stiglitz version of the model in the late 1970s. Especially suspect is the symmetry assumption that all versions of a single differentiated product are equal substitutes for all others, instead of some being close substitutes (e.g., different makes of small cars) while others are more distant substitutes (e.g., small cars and luxury cars).²⁹ The model also uses a single representative consumer who consumes some of all variants of each differentiated product. Aggregation to the whole economy is then done trivially—simply by multiplying the representative consumer by n . But this representative consumer cannot be a “real representative person” since no one person consumes all variants of all differentiated products. Instead the “individual” must himself be an aggregation from actual individual consumers, each one of whom consumes only a small fraction of all the available variants of each generic commodity. However, no one knows how to do such an aggregation. Certainly, it cannot be done when individual consumer's tastes are specified in characteristic space, which is the natural way to model the empirically established consumer behaviour with respect to any differentiated

commodity whose variants individual consumers perceive as closer and more distant substitutes for each other.³⁰

Aggregation and end-state competition

Furthermore, formal aggregation requires the existence of the *end state* of perfect or monopolistic competition, rather than existence of the *process* of competition that we actually see in the real world. This important point requires elaboration. Standard neoclassical theory treats competition as an *end state* that is perfectly competitive equilibrium. This stands in contrast to the view usually associated with the Austrian tradition, which sees competition as a *process* that takes place in real time (Blaug 1997). In the latter:

“...firms jostle for advantage by price and non-price competition, undercutting and outbidding rivals in the market-place by advertising outlays and promotional expenses, launching new differentiated products, new technical processes, new methods of marketing and new organisational forms, and even new reward structures for their employees, all for the sake of head-start profits that they know will soon be eroded. ...[in short] competition is an active process.” (Blaug, 1977: 255-6)

All of the aggregations that underlie the theories that use the neoclassical aggregate production function assume the existence of end-state competition because the aggregations cannot be made formally given ongoing *process competition*. Clearly, however, it is process not end state competition that we see in the real world. The judgement of economists varies greatly on how much this matters. The standard treatment, however, is an uneasy combination of insistence on a high level of rigour, where that is possible, and the application of the intuitive judgement that deviations caused by non-fulfilment are quantitatively unimportant where the rigorous conditions cannot be fulfilled (which when aggregating production functions is more or less everywhere).³¹

Appreciative aggregation

Thus there is little, if any, correspondence between what is observed technological change and the behaviour of firms at the micro-level and the aggregate data and models that economists use to explain growth at the macro level. This ‘aggregation problem’ remains an almost totally ignored issue: *how can we build a bridge designed to aggregate the kind of innovative behaviour that we observe at the micro economic level to the macro behaviour that economists explain by neo-classical-type growth models that use a single aggregate production function.*

So we need to use what Richard Nelson calls appreciative theorising to make the bridge between micro observations that we see as relevant to growth and a macro model that captures this behaviour. The task then is to make macro models that come as close as possible to mirroring the real behaviour we see at the micro level even though we cannot aggregate from the one to the other. How to do this properly requires a much deeper

analysis than we have space for here. In the meantime, we develop a macro model that contains some of the key characteristics that we see at the micro level—even though there is no known way to aggregate from our micro observations to the macro model that we use even though we do use many of the standard tools, and are thus guilty of making many of the standard simplifications that go into this type of macro growth modelling. We concentrate on process competition by not imposing equilibrium requirements on the model and we allow for uncertainty in most of our relations. We treat the creation of technological knowledge as an endogenous process and we explicitly model some of the complementary relationships of knowledge production, separating technological externalities and complementarities specifically. We also generate a process of continual growth that does not settle into an equilibrium steady state.

III. A NON-EQUILIBRIUM MODEL OF SUSTAINED GROWTH

We have developed a view of the growth process that is consistent with the micro observations of technological change. To summarise: new GPTs arrive at widely dispersed times (usually there is more than one GPT working through the system at any one time). They rejuvenate the growth process by presenting a new research program for applied R&D. The evolution of the GPT and its new derivative technologies create a growing set of technological complementarities. Technological changes enabled by the new GPT, in the form of new technologies and improvements in existing technologies, are slow at first, then accelerate, and finally slow again as the potential of the GPT is more fully exploited. As GPTs follow each other, each one brings its own research program whose results follow its own logistic curve. Although some programs are richer than others, there is *no* expectation that successive GPTs will always either accelerate or decelerate growth. But they do sustain it. As long as GPTs continue to arrive on the scene, growth will continue—sometimes faster and sometimes slower.

In this section, we outline a simple growth model that captures some of these key characteristics of technological change. At this stage, we omit the logistic curve approximation to the GPTs development trajectory (although we intend to use it in later versions of the model.). This allows us to avoid introducing increasing returns into any of the model's stages of production, but at the cost of not being able to explain the productivity slowdown that seems to accompany the first stages of a new, pervasive GPT. Our motivation is to understand the growth process and the technological complementarities that drive it. To this end, we wish to show that growth does not have to exhibit the long run stationary equilibrium properties of the standard AK models nor does sustained growth require increasing returns or externalities. (For an overview of AK style models see Romer 1996 chapters 3 and 4.) Although our analysis does not preclude the existence of any of these, it shows that growth can be sustained even if they are small or non-existent.

Ours is a three-sector economy, with each sector represented by a single production function.³² Technological complementarities among different types of R&D are modelled explicitly. The sector “output” produces consumer's goods, which we assume can be aggregated into the single output index, C_t ; the sector “applied R&D” does R&D that applies the general purpose technology to specific uses, the sector “general purpose R&D” does fundamental R&D that generates new GPTs. Each of these activities uses a single generic input (resources) to produce

their respective outputs in a production function that displays decreasing returns to resource inputs.

A Two Sector-Model

We first introduce a simple two-sector model to act as a benchmark.

Input: The resources devoted to the production of “consumer’s goods”, “applied R&D”, and “general purpose R&D” are respectively $r_{c,t}$, $r_{a,t}$ and $r_{g,t}$. (In the two-sector model $r_{g,t} = 0$.) The resource constraint for the economy is:

$$[\text{III.1}] \quad R_t = r_{c,t} + r_{a,t} + r_{g,t}$$

where R represents all of the factors of production that are not technology (i.e., endowments given by the natural world including land, minerals and raw labour, and capital both physical and human).³³ So the usual growth variables are contained in R and we could include a standard factor accumulation procedure in our modelling framework. We could employ Solow’s assumption of constant returns to inputs (R for us) but for our purposes it is easier to assume decreasing returns to R for what follows. We justify this by observing that some of the variables in R are fixed (e.g., population in industrialised countries and natural resources everywhere) so that those that can be accumulated will be subject to decreasing returns.³⁴ This is conceptually identical to Solow’s model where labour is exogenous. An accumulation exercise for R (subject to decreasing returns) or a subset of R (subject to decreasing returns) arrives at the same result as in the Solow model with labour constant. Endogenous factor accumulation results in increases of R up to some steady state value. So, allowing for saving and accumulation in our framework produces the result that factor accumulation with labour constant and without technological change runs into decreasing returns. For simplicity we do not model factor accumulation, so that R is held constant over time.

Output: Consumer’s goods are produced by an aggregate production function that uses resources, $r_{c,t}$.

$$[\text{III. 2}] \quad C_t = \mu A_t r_{c,t}^\alpha$$

where $0 < \alpha < 1$. A_t , is a productivity parameter that is determined by the accumulated knowledge produced in the applied R&D sector. Increases in A_t cause increases in the marginal productivity of resources in consumer’s goods. In the two sector model μ is constant at a value of unity.

Applied R&D: This sector creates capital for the output sector with a production function that uses resources, $r_{a,t}$, as inputs. B_t^a , is a productivity parameter that is determined by the accumulated knowledge produced in the GPT sector.

$$[\text{3}] \quad \begin{aligned} a_t &= \nu B_t^a r_{a,t}^\beta \\ A_t &= a_t + (1 - \varepsilon) A_{t-1} \end{aligned}$$

where $0 < \beta < 1$, ν is a parameter used in the simulation to calibrate the model and ε is the rate of obsolescence that applies to the stock of R&D-produced knowledge. Increases in B_t^a increase the marginal productivity of resources in applied R&D and this in turn affects the marginal productivity of resources in output.

In this simple version, resources devoted to R&D produce disembodied technological change. For the more realistic case where technological change is embodied in new capital, gross investment can be in capital that embodies the new technologies (while net investment is zero). (This amounts in our model to holding R constant while A and B become the indexes of the quality of new capital.)

Simulation of the two-sector model

If we take B_t^a as a constant parameter, we have a standard AK style growth model that exhibits steady state properties. In the absence of either exogenous growth of the parameter B_t^a or of the resources, R_t , there will be no growth after the system settles into its steady state.

Resources are allocated to maximise total output defined as the sum of the expected discounted value of goods production over the infinite horizon

$$\begin{aligned}
 & \max_{a,r} \sum_{t=0}^{\infty} \rho^t (C_t) \\
 & \text{s.t.} \\
 & A_t = a_t + (1 - \varepsilon)A_{t-1}
 \end{aligned}$$

[4]

where ρ is the rate of time discount. There is no notion of long run equilibrium imposed on the system, nor do we solve for one. Instead resources are allocated to the two sectors in order to maximise the present value of total output as if by perfectly competitive agents in the two sectors.³⁵ Thus, agents in the consumer's goods sector pay the opportunity cost of A_t in terms of foregone resources that go into the R&D sector. The marginal products of resources in the consumer's goods and applied R&D sectors are constantly in transition and are determined by the amount of resources devoted to the two sector and by A_t and B_t^a . However, growth must eventually stop. This result comes out of a dynamic optimisation exercise and occurs because as time passes and A_t is increased by the activities of the applied R&D sector, the marginal productivity in direct production of C rises steadily, causing fewer and fewer resources to be allocated to producing C indirectly through applied R&D. Eventually, A_t becomes so large that the resources devoted to R&D are just enough to compensate for the depreciation of technological knowledge. At this point, there is no further growth. (See Romer, 1996, chapters three and four.)

We use Microsoft Excel to perform our numerical simulations.³⁶ The parameterisation for the two sector model is as follows:

$$\alpha = 0.1$$

$$\beta = 0.1$$

$$\nu = 0.1$$

$$\mu = 1$$

$$A_0 = 1$$

$$R = r_{c,t} + r_{a,t} = 1000$$

$$\varepsilon = 0.02$$

B^a is 1 initially and then changes exogenously to 10 in period 96.

Figure 2³⁷ shows the plot of the value of the output from the consumer's goods sector on the first 196 time periods.³⁸ In the ninety-sixth period the change in the parameter B is introduced. This rejuvenates the growth process. There is an increase in the allocation of resources to the R&D sector (whereas up to then the resources in the R&D sector had been falling).

Figure 3 shows the plots of the optimal allocations of resources in the goods and the R&D sector for each of the 196 periods. As productivity grows in the consumption sector, resources fall in the R&D sector and rise in the consumption sector. Then the exogenous increase in B_t^a in period 96 increases the marginal productivity of resources devoted to R&D and this causes a diversion of resources out of goods production into R&D. After this resources are slowly transferred back in C, once again slowing the growth process.

A Three-Sector Model

We now introduce a further, general purpose, layer of R&D activity that endogenises the B_t^a term. We introduce randomness into this version of the model in three ways. First, the amount of new knowledge generated by resources devoted to general purpose R&D in every period is uncertain. Second, the point in time when a new GPT arrives is uncertain. Third, the affect that the GPT has on productivity in the R&D sector when it does arrive is uncertain. These characteristics are designed to conform with the empirical evidence that the search for new fundamental technologies is an uncertain venture in three senses that (i) one cannot know how much knowledge will be generated by a given amount of effort (ii) success occurs only occasionally and (iii) the applied potential of GPTs varies from one to another.

General purpose R&D: This sector produces general purpose knowledge, b_t , using a production function that has decreasing returns to the resources it employs, $r_{g,t}$, and a random component to reflect the fact that one is never sure how much knowledge will be generated by a given amount of resources allocated to fundamental research.

General purpose knowledge accumulates as a stock, B_t^p . Although it is potentially useful, new stocks only become practically useful to the applied R&D sector when a breakthrough occurs to invent a new GPT. This occurs as a result of general purpose R&D but at randomly determined times. When it does occur, the value to the applied R&D sector, B_t^a , depends on how much general purpose knowledge has been accumulated since the last GPT and a random variable that reflects the fact that the applied potential of GPTs vary in ways that cannot be predicted when they are originally being developed.

We also need to model an important feedback from the applied R&D sector to general purpose R&D. In practice, knowledge produced in the applied sector is potentially useful to both the consumers' goods sector and the general purpose sector. Examples include.

- Torricelli's demonstration of the weight of air in the atmosphere was of utmost scientific importance and came out of his attempts to design an improved pump.
- Carnot's creation of the science of thermodynamics was an outgrowth of attempts to explain what determined the efficiency of steam engines.
- Joule's discovery of the law of conservation of energy came from work on alternative sources of power for his father's brewery.
- Pasteur's development of the science of bacteriology came out of his studies of fermentation and putrefaction in the French wine industry.

We could model these as unintended spillovers from applied to general purpose R&D. Instead we model agents as competitively dividing applied knowledge between that part that is used to increase productivity in production of the consumers good and the rest that is used to increase productivity in general purpose research. We make this assumption for two reasons. First, we want to ensure that we do not introduce externalities or increasing returns to scale into the model. Second, this set of complementarities is relatively simple to formulate in our framework.

Equation [III.5] defines the production function for general purpose knowledge. It introduces the additional problem of the allocation of applied knowledge between increasing efficiency in the production of consumption goods and in general purpose R&D. μ , which is the proportion of applied R&D knowledge allocated to the consumer's goods sector, is chosen to maximise the objective function defined in equation [III.10] below. We assume that agents form a subjective expectation of the consumption value of general purpose R&D (which they must do if they are to take decisions under uncertainty rather than risk). So μ is chosen to maximise expected returns from the general purpose sector and the consumption output. The residual $(1 - \mu)$ is allocated to the general purpose sector.³⁹

The stock of potentially useful general purpose knowledge is defined in equation [III.6] below. General purpose knowledge accumulates because of the output from b_t . In each period, some of the stock of general purpose knowledge is rendered obsolete and thus the stock is reduced by the obsolescence factor δ .⁴⁰

The accumulation of knowledge in the general purpose sector only occasionally spills over to be useful to the applied R&D sector. This happens when a new GPT is invented, which in our model is when a random variable, λ , exceeds some threshold value, as shown in equation [III.7].

Let the general knowledge production function be

$$[\text{III.5}] \quad b_t(r_{g,t}) = (1 - \mu)A_{t-1}\theta_t r_{g,t}^\sigma$$

where $0 < \sigma < 1$, and θ is a random variable.⁴¹

Let the accumulation of potentially useful general purpose knowledge be:

$$[\text{III.6}] \quad B_t^p = b_t(r_{g,t}) + (1 - \delta)B_{t-1}^p$$

Let the stock of general purpose knowledge that is useful for applied R&D at any point in time be:

$$[\text{III.7}] \quad B_t^a = \begin{cases} B_{t-1}^a + \vartheta_t (B_t^p - B_{t-1}^a) & \text{if } \lambda \geq \lambda^* \\ B_{t-1}^a & \text{otherwise} \end{cases}$$

where λ^* is the threshold value of λ and ϑ is a random number that takes on only positive values (many of which can be fractions). ϑ determines the effect on R&D of the knowledge accumulated in the GPT sector since the last GPT occurred.

We choose λ and ϑ from skewed distributions and θ from a uniform distribution with a mean of 1 so that on average the resources devoted to the general purpose sector will be as effective as resources elsewhere in the system for all of the simulation runs reported below.⁴² In specific realisations θ will be more or less productive than this average. The uniform distribution on $[0,a]$ generates random numbers according to its cumulative density function CDF.

$$[\text{III.8}] \quad x = G(z) = \frac{z}{a}$$

In order to introduce skewness into the distribution we define a random number as being generated from the following cdf:

$$[\text{III.9}] \quad x = F(z) = \left(\frac{z}{a}\right)^n$$

which is achieved by raising the random number generated from the uniform distribution to some power n . This skews the distribution to the right as indicated by the fact that the mean is greater than the median (Pfaffenberger and Patterson, 1987, 42-3).⁴³

We choose $a = 1$ and $n > 1$ so that the concentration of mass in the probability density function is skewed toward zero.⁴⁴ We do this for both λ and ϑ in order to reduce the likelihood and probable size of impact of new GPT so that we can space out their timing and affect. For example, in generating values for λ , this lowers the probability that λ surpasses the threshold value in any one period. If the threshold value of λ were 0.5, then the random number generated from the

uniform distribution with $a = 1$ would achieve the threshold with probability 0.5, whereas the random number generated from the skewed distribution with $a = 1$ and $n > 1$ will surpass the threshold value with a lower probability. The larger is n the smaller is the probability that $\lambda > \lambda^*$.

We can also scale up the magnitude of random numbers generated by multiplying the number drawn from either the uniform or the skewed distribution by a scalar, s . This is done for example with ϑ in order to increase the variance of affect that GPTs can have on the applied R&D sector when they arrive. So we have,

$$[\text{III.10}] \quad \vartheta = s \left(\frac{z}{a} \right)^n$$

Thus, while the probability of getting a large impact from a GPT is low due the skew ness of the distribution, the possible impact can be large, depending on the size of s .

It remains to show what determines the amount of resources, $r_{g,t}$, that will be devoted to pure GPT research. Because there is uncertainty about the opportunity cost margins for resources being allocated to the general purpose sector and the other sectors, we must make some assumptions about the objective function for the maximisation. We maintain the assumption implicit in the previous model that utility is monotonically increasing in consumption.⁴⁵ But we must now make a further assumption about how agents form expectations of the future consumption payoff from resources devoted to general purpose R&D. The normal margins on which optimal resource allocations get carried out do not exist here because of the uncertainty associated with the output of general purpose R&D. Possible assumptions include agents being endowed with natural curiosity, deriving positive utility from the pursuit of general knowledge; alternatively, agents might form subjective expectations of the consumption values of using resources to produce knowledge based on the history of past knowledge breakthroughs. In reality both assumptions are probably partly correct.

In our model, these expectation might be based in part on past observations of λ , θ and ϑ .⁴⁶ However, for simplicity in our first round of modelling, we assume that these expectations of the consumption or utility value of resources devoted to general purpose R&D are expressed by the variable Φ in the maximisation of [III.11] below, and we make it a constant.

In the three-sector model, agents allocate resources so as to maximise discounted value of consumption and production in the general purpose R&D sector measured in expected consumption units over the infinite horizon:

$$[\text{III.11}] \quad \begin{aligned} & \max_{\{c, r, g, \mu\}} \sum_{t=0}^{\infty} \rho^t (C_t + \Phi b_t) \\ & s.t. \\ & A_t = a_t + (1 - \varepsilon) A_{t-1} \\ & a_t = \nu B_t^a r_{a,t}^\beta \end{aligned}$$

$$\begin{aligned}
B_t^p &= b_t(g_t : \theta) + (1 - \delta)B_{t-1}^p \\
b_t &= (1 - \mu_t)A_{t-1}\theta_t r_{g,t}^\sigma \\
B_t^a &= \begin{cases} B_{t-1}^a + \vartheta(B_t^p - B_{t-1}^a) & \text{if } \lambda \geq \lambda^* \\ B_{t-1}^a & \text{otherwise} \end{cases}
\end{aligned}$$

where again $\rho < 1$ is the rate of time discount. Adding b and C is possible if b is measured as the forgone consumption as a result of allocating resources to pure R&D rather than the production of consumption goods.

Again we employ Solver iteratively so that allocating resources to each of the three sectors maximises total output in each period. A_t is allocated to the consumption and general purpose sector by choosing μ . The production resulting from the resource allocations in the applied R&D and GPT sectors are then fed into the appropriate terms (A_t and B_t) and the maximisation exercise with respect to resources is repeated in the next period, and so on.

Simulation in a three sector model

With B_t^a endogenously determined, we now simulate the growth process for the three sectors of the economy. The parameterisation for this baseline simulation is as follows:

$$\alpha = 0.1$$

$$\beta = 0.1$$

$$\sigma = 0.1$$

$$\nu = 0.1$$

$$A_0 = 1$$

$$B_0 = 1$$

$$R = o_t + r_t + g_t = 1000$$

$$\varepsilon = 0.02$$

$$\delta = 0.0001$$

$$\Phi = 1$$

We chose θ from a uniform distribution on the interval $[.8, 1.2]$ and ϑ from skewed distribution with $a = 1$, $n = 5$ and $s = 4$. We choose λ from a skewed distribution on $[0,1]$ with $n = 15$ and $s = 1$. Part a of Figure 4 shows the time series of the output from the consumers' sectors and the stocks of applied and general purpose knowledge. Part b shows the growth rate of the consumers sector. Figure 5 shows the plots of the optimal allocations of resources in the goods and the R&D sector for each of the 300 periods that the model was allowed to run.

When a new GPT arrives, there is an increase in the marginal productivity of resources devoted to the applied R&D sector and thus there is a diversion of resources out of goods production into R&D production. The simulation results are similar to the two sector model, except that GPTs are now determined endogenously. The arrival of a new GPT makes applied R&D more productive. R&D rises, increasing the growth in productivity in the consumption good and so causing the growth of output to increase. The growth rate then falls steadily until another GPT arrives, again rejuvenating the growth process.

Is this GPT-driven growth sustainable in the long run? The most straight forward way to show that growth is sustained in our model is to turn off the uncertainty. We let λ , θ and ϑ be one in every period. This is equivalent to having the resources that are devoted to general purpose R&D yield a return solely determined by the production function, having GPTs arrive every period with certainty, and having a uniform impact. When we run the simulation, the system converges to a constant growth rate where consumption, the stock applied R&D knowledge and the stock of general purpose knowledge are all growing at the same rate. These results are shown in Figure 6.

Sustainable growth with decreasing returns to scale

To emphasize that without GPTs the productivity of R&D will fall off, we use a slightly different formulation of the model that produces sustained growth driven by technological complementarity. In this version, we assume that the production functions in each sector have constant returns over the combination of resources and accumulated knowledge. Thus, the production functions exhibit decreasing returns to ceteris paribus increases in technological knowledge. However, we add the assumption that current GPTs spillover into future GPTs by increasing the potential impact of future GPTs. To do this we endogenize the parameter s from equation [III.10] used to scale up ϑ . Now the arrival of a GPT will increase the possible size of future GPTs because the term multiplying the skewed distribution that ϑ is drawn from is growing due to GPTs.

$$[\text{III.12}] \quad s(B_t^a) = x(B_t^a)^\omega$$

which we set equal to a linear function of the accumulated knowledge from the general purpose sector by setting $\omega = 1$ with x being a constant parameter. We apply $s(B_t^a)$ to the random number ϑ .

We alter the production functions as follows:

$$[\text{III.2}'] \quad C_t = (\mu A_t)^{\alpha_1} r_{c,t}^{\alpha_2}, \quad \alpha_1 + \alpha_2 = 1$$

$$[\text{III.3}'] \quad a_t = \nu (B_t^a)^{\beta_1} r_{a,t}^{\beta_2}, \quad \beta_1 + \beta_2 = 1$$

$$[\text{III.5}'] \quad b_t(r_{g,t}) = ((1-\mu)A_{t-1})^{\sigma_1} \theta_t r_{g,t}^{\sigma_2}, \quad \sigma_1 + \sigma_2 = 1$$

Once again we run the numerical simulation, using Solver to allocate resources among the sectors to maximize equation [III.11]. The parameter values are as follows:

$$\alpha_1 = 0.4$$

$$\alpha_2 = 0.5$$

$$\beta_1 = 0.25$$

$$\beta_2 = 0.4$$

$$\sigma_1 = 0.25$$

$$\sigma_2 = 0.5$$

$$\nu = 0.1$$

$$A_0 = 1$$

$$B_0 = 1$$

$$R = o_t + r_t + g_t = 1000$$

$$\varepsilon = 0.02$$

$$\delta = 0.0001$$

$$\Phi = 1$$

To simplify we chose θ to be constant and equal to one. ϑ is chosen from skewed distribution with $a = 1$, $n = 5$ and multiplied by $x B_t^a$ where $x = 1$. We choose λ from a skewed distribution on $[0,1]$ with $n = 15$.

Part a of figure 7 shows the plot of the consumers' goods and total output on time and part b shows their growth rates.

[Insert figure 7]

The spikes in the plot of the growth rate correspond to the introduction of new GPTs. There is no pattern in the terms of the size of impact that one GPT has relative to another and large impacts may occur after small impacts, and vice versa, according to the realization of ϑ . The important feature of the growth pattern is that GPTs rejuvenate the growth process and there is no trend towards ever-smaller impacts. If the spillover from current to future GPTs was not present, the growth rejuvenation resulting from each subsequent GPT would be smaller and smaller. The growth process is sustainable. This is verified by removing the uncertainty setting λ and ϑ equal to one in each period. The system exhibits unbounded growth in this case as the spillover from of the GPT onto itself implies that there are increasing returns to scale in the general purpose knowledge sector.

This version of the model shows that we can capture more of the stylized facts presented in Part I by employing a model where we have DRS to accumulating knowledge (i.e., no sustained growth by itself) but with a spillover from one GPT to the next. Having decreasing returns to the accumulated knowledge implies that the applied research trajectory kicked off by the arrival of a new GPT will eventually run into decreasing

returns to R&D investment. The spillover from the GPT sector onto the future potential for GPTs allows us to capture the fact that GPTs open up possibilities for further discoveries but do not guarantee further discoveries will be made or that they will payoff significantly. This then allows us to differentiate standard externalities from some of the spillovers included in the set of technological complementarities.

To see this suppose there were certainty in the model. Then the spillover would be a standard externality. There would be a continual activity conferring higher than normal returns in every period that would show up as increasing returns to general purpose knowledge. However uncertainty makes this spillover something different. It is a spillover that is not a standard externality. Instead when the GPT arrives it creates opportunities for future GPTs to have bigger impacts on the economy. This is consistent with the facts presented in section 1 where we argued that GPTs open up opportunity sets for exploration and some GPTs enable others.

IV. CONCLUSION: SUSTAINED GROWTH AND ALL THAT

Capitalist economies have produced continued growth in per capita income at least since the end of the eighteenth century. This growth has not been at a constant rate. For example, the measured growth rate⁴⁷ in the United States was exceptionally high in the period following the Second World War until the early 1970s, then fell over the next two decades, accelerating again sometime in the early 1990s.

The literature contains many theoretical models that produce sustained growth. All of them, however, produce either constant growth along a steady state growth path or continually accelerating or decelerating growth rates. None explain the variable growth rates that we observe. We briefly survey the most important of these in order to put our paper into perspective.

A Solow growth model with labour augmenting, Harrod neutral, technical change produces steady state growth at the exogenously determined rate of technical change. There is overwhelming evidence that technical change is endogenous⁴⁸ but no evidence that all technical change is always labour augmenting.⁴⁹ An alternative to this is to assume that exogenous technological change augments both labour and capital which yields what is termed Hicks neutral or disembodied technological change. This assumption is seldom used in theoretical work because it does not produce a steady state equilibrium, although there is no evidence that would suggest it is less in line with real behaviour than the assumption of labour augmenting technical change. In either case the explanation of growth in per capita income is that it is exogenous.

Romer's 1986 model produces sustained growth through an externality in investment where each firm's productivity growth is a result of its own investment and total national investment. The model implies aggregate increasing returns to investment, for which no compelling empirical evidence has been found, and a constantly accelerating growth rate, which is clearly counterfactual.

Lucas's model introduced human capital into the production function and produces sustained balanced growth if the production function displays constant returns to the accumulating factors, including human capital. Lucas' goods output equation is:

$$Y = AK(t)^\beta [\mu(t)h(t)N(t)]^{1-\beta} h_a(t)^\gamma$$

which is divided between consumption and investment in physical capital. Y represents goods, K is physical capital, N is population and A is a parameter. μ is the proportion of human capital devoted to the production of goods and $(1 - \mu)$ is devoted to the production of human capital. He has a human capital producing sector that uses only current human capital. $\dot{h}(t) = \delta(1 - \mu)h(t)$ which is linear in $h(t)$. There is an externality in the model denoted by $h_a(t)$ which represents external human capital that contributes to goods production (Lucas (1988)). The important feature of Lucas model is that both the goods sector and the human capital sector exhibit at least constant returns to the accumulating factors of physical and human capital.

There is also the class of research and development models that David Romer (1996) attributes to Paul Romer (1990), Grossman and Helpman (1991a) and Aghion and Howitt (1992) and which come closer than the other models to our formal framework. These are two sector models like Lucas's, where changes in the stock of technological knowledge, A , are produced in one sector with part of the supply of labour and capital, and with A itself. Goods are produced with the residual of labour and capital, and A augments labour as in the Solow model. So in this framework, A gets double counted as an input into goods production and an input into its own production. Growth is sustainable provided the returns to capital and the stock of technological knowledge are at least constant.

Along with the models that seek to measure externalities by associating R&D with total factor productivity at various levels, all of the endogenous models to the extent that the deal with technological knowledge treat it as an input into the production function rather as something that shifts that function. This is no trivial matter.

The standard microeconomic treatment defines the production function in terms of traditional inputs, which in the most aggregated case, gives:

$$(III.13) \quad Y = Y_1(L, K)$$

and in the simple Cobb Douglas example:

$$(III.14) \quad Y = AL^\alpha K^\beta \quad \alpha + \beta = 1$$

Constant returns to scale are assumed on the empirical grounds that replication of identical productive facilities should allow output to grow in proportion to inputs (at least for integer multiples of the inputs used in the production facility in question). Changes in technology brought about by R&D are assumed to alter the nature of the production function, or the efficiency with which one of the factors is used. In the latter case, the input alters when it is measured in efficiency units. R&D is *not* assumed to be an input itself.

The models that treat knowledge as an input are in striking contrast. In them:

$$(III.15) \quad Y = Y_2(L, K, D)$$

and in Cobb Douglas form:

$$(III.16) \quad Y = AL^{\alpha}K^{\beta}D^{\gamma}$$

where D is, depending on the application, the flow of current R&D expenditure, the stock of accumulated R&D, or human capital. This formulation has several problems. First, it implies that at the micro level, a firm that wishes to replicate its production facilities to increase its output with constant technology must alter its R&D in proportion to its conventional inputs. Given that technology is assumed to be constant, this is an undesirable implication. Second, there is no known behaviour, similar to that of replication, that would predict constant returns to all traditional inputs plus R&D. Constant returns in [III.16] must then be assumed to be a fortuitous accident that we have no reason to believe would be a universal phenomenon.

If D is taken to mean human capital in [III.16], then it might be assumed that this includes technological knowledge, which accumulates along with other constituents of D . But such formulations are often used empirically to determine the amount of growth that is due to the accumulation of factors on the one hand and to technological change on the other. If we are to deal with this question we must know what it means to hold technology constant while physical and human is accumulating. To do this, we would need to do the following: hold all product, process, and organizational technologies constant at what was known at some base period; accumulate more physical capital that embodies the technologies known at the time; accumulate more human capital in the form of more education *only in what was known at that time*. So if human capital includes new technological knowledge that is put to use, then much applied technological change is included in this variable and a fitting of equation [III.16] to the data cannot be used as evidence on how much growth is the result of factor accumulation and how much of technological change. In short, if we wish to determine how much growth is due to the accumulation of capital and how much to technological change we cannot define capital so that it includes technological change⁵⁰

In our models, we formulate the production function in terms of the traditional inputs that would need to be altered when production facilities were replicated.⁵¹ We then model technological knowledge as altering that function. But what assumptions should be made with respect to technological knowledge?

Much attention is paid in the literature as to whether the accumulation of knowledge has constant or diminishing returns. In the formulations such as (III.15) and (III.16) if the overall function has constant returns to scale, then *ceteris paribus* accumulation technological knowledge displays decreasing returns (assuming that is what D stands for, either in the guise of human capital or the stock of R&D). In the first of our two 3-sector models, there are constant returns to accumulating knowledge, in both the applied and the GPT sectors. This is seen by the coefficient of unity on A in equations [III.2] and [III.5],

and on B^a in equation [III.3]. So a key part of this model is that ceteris paribus increases in A have a constant effect on efficiency in the production of C (equation [III.2]) and of GPT knowledge (equation [III.5]), while ceteris paribus increases in B^a have a constant effect on the efficiency of the applied R&D sector (equation [III.3]). This is sufficient for sustained growth in this model.

However, the growth is also dependent on the key assumption that increases in applied knowledge feed back to increase the efficiency of the GPT research sector (equation [III.5]). Without this technological complementarity between the two research sectors, growth would peter out. The relative efficiency of direct production of C in equation [III.2] would grow without limit while the efficiency of research would not so. Over time, more and more resources would be pulled into direct production until eventually growth would stop. What stops this happening is a technological complementarity⁵² with no externality since the GPT sector must still cover its direct research costs and must pay those in the A sector for the extra efficiency that A 's R&D confers on B . Furthermore, this is not the kind of increasing returns that is associated with an investment externality, as in Paul Romer's early models. Instead, it merely says that the research program of each research sector impacts favourably on the efficiency of the other research sector. Our modelling of this technological complementarity between the two R&D sectors provides a fundamental distinction from David Romer's (1996) research and development models. We distinguish different types of knowledge production, allowing us to model technological complementarities explicitly, and we use the output of one knowledge sector as a state of technology index in other sectors, not as an input to them. Furthermore, we do not double count the knowledge produced by using it as an input into two sectors.

The second of our two 3-sector models reinforces these results. Here we assume decreasing returns to knowledge by making the production functions in [III.2] [III.3] and [III.5] constant returns in resource inputs and knowledge, so that we have diminishing returns to the accumulating factor, knowledge. Although, as argued above, we do not think this type of formulation (with knowledge in the production function rather than shifting it) captures reality, it helps us to emphasise the importance of technological complementarities. In this model, as well as having technological complementarities that spill over from one research sector to the other, we allow for a temporal complementarity, of the type also discussed in Part 1. What is learned in the GPT sector today makes future GPT research more efficient. This also give rise to sustained growth.

Finally, we must ask if all we have done is to smuggle in increasing returns by another door? Our object, consistent with Nelson's concept of appreciative theory, is to build a model that is consistent with the facts outlined in Part 1. One fact that is clear is that R&D does not encounter constant returns over a long time period in the sense that as time passes the same allocation of resources produces the same absolute contribution to increasing output, measured, say, in constant dollars. If that was so, R&D would be becoming less and less efficient than the production sector where the same allocation of resources is producing more and more over time. A neutral assumption would be that the

efficiency of the R&D sectors would increase at the same rate as the efficiency of the sectors that produced final output. Thus if, say 2% of the nation's resources were allocated to R&D over a long period and they produced an increase in productivity, this would be at constant *rate* rather than a constant *absolute amount* (and hence a declining rate of growth of productivity). This is the assumption that is made in all applied studies of R&D, and it is what the evidence supports. Otherwise, as indirect production through R&D became less efficient relative to direct production of final goods, the allocation of resources to R&D would shrink over time. So the evidence suggests that the absolute ability of R&D to add to output is rising over time sufficiently to keep its contribution to the growth rate constant. All we have done is to model such a process rather than assume it.

In the process we have built a model with of number of characteristics that seem to mirror what we observe about actual growth in several important ways

- The model has sustained growth generated by the arrival of successive GPTs.
- GPTs rejuvenate the growth process by making applied R&D more productive while applied and GPT R&D increase the efficiency of GPT R&D over time. These relations mirror the types of technological complementarities that are observed empirically, and were discussed in Part 1.
- The model does not have a steady state growth path. Instead, a very productive GPT can accelerate the growth rate while a less productive GPT can slow it. (In contrast the models by Helpman and Trajtenberg (1998a), Helpman and Trajtenberg (1998b) and Aghion and Howitt (1998) all have each successive GPT being identical to the preceding one.
- Our GPTs are the result of endogenous R&D while the the models referred to above make them exogenous.
- Although endogenous, the arrival of our GPTs is subject to random variation. Both Helpman and Trajtenberg models have them arriving at fixed intervals while Aghion and Howitt's arrive at randomly determined times.
- Our models do not require the type of ad hoc assumptions used by other modellers. For example, Aghion and Howitt (1998) need to assume that the amount research in each sector is determined by a fixed endowment of specialized labour that is unemployed except when newly arrived GPTs are being readied for application to production.

In summary, ours is a model of GPT generated, sustained growth operating through technological complementarities with neither increasing returns nor conventionally defined externalities.

BIBLIOGRAPHY

- Aghion, P. and P. Howitt (1992), "A Model of Growth Through Creative Destruction", *Econometrica*, 60, 232-396.
- _____ (1998) "On the Macroeconomic Effects of Major Technological Change", in (E. Helpman (ed) *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press), 121-144.
- Arrow, K. J. (1962), "Economic welfare and the allocation of resources for invention". In *The rate and direction of inventive activity: Economic and social factors*, 609-25, NBER Special Conference Series vol. 13. Princeton, N.J.: Princeton University.
- Blaug M.(1997), "Competition as an end-state and competition as a process" in *Trade Technology and Economics: essays in honour of Richard G. Lipsey.*, B Curtis Eaton and Richard G. Harris (eds.) (Cheltenham: Edward Elgar)
- Chandler, A. D. (1977), *The Visible Hand: The Managerial Revolution in American Business*, (Cambridge, Mass.: Belknap Press).
- Eaton B. C., and R. G. Lipsey (1989), "Product Differentiation", in *Handbook of Industrial Organisation Volume One*, (R. Schmalensee and R.D. Willig eds.) (Amsterdam: North Holland)
- _____ (1997), *On the Foundations of Monopolistic Competition and Economic Geography* (Cheltenham: Edward Elgar).
- Freeman and Louca, (2001), *A Time Goes By: From the Industrial Revolutions to the Information Revolutions*, (Oxford: Oxford University Press)
- Griliches, Z. (1980), "Returns to Research and Development Expenditures in the Private Sector", in *New Developments in Productivity Analysis*, W. Kendrick and B. N. Vaccara (eds) (Chicago: University of Chicago Press).
- _____ (1982), "The Search for R&D Spillovers", *Scandinavian Journal of Economics*, 94 supplement, S29-S47.
- _____ (1984), "Productivity, R&D, and the Data Constraint", *American Economic Review*, 84(1), 1-23.
- Grossman. G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, (Cambridge, Mass.: MIT Press).
- Helpman, E. and M. Trajtenberg (1998a) "A Time to Sow and a Time to Reap" in (E. Helpman (ed) *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press), 55-84.
- _____ (1998a) "Diffusion of General Purpose Technologies" in (E. Helpman (ed) *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press), 85-120.

- Jorgenson, D. and Z. Griliches (1967), "The Explanation of Productivity Change", *Review of Economic Studies*, 34(3) 249-83.
- Lipsey, R. G. (2000), "Sources of Continued Long-run Economic Dynamism" in *The Future of the Global Economy: Towards a Long Boom?* Wolfgang Michalski, Riel Miller and Barrie Stevans eds., (Paris: OECD)
- _____ (2001, forthcoming), "Successes And Failures In The Transformation Of Economics", *The Journal of Economic Methodology*.
- _____ and Bekar (1995), "A Structuralist View of Technical Change and Economic Growth" in *Bell Canada Papers on Economic and Public Policy* Vol. 3, Proceedings of the Bell Canada Conference at Queen's University, November 1994, (Kingston: John Deutsch Institute).
- _____ and Carlaw, (2001, Forthcoming), "What Does Total Factor Productivity Measure?" forthcoming as a Statistics Canada technical paper.
- _____ C. Bekar and K. I. Carlaw (1998a), "What Requires Explanation?" in E. Helpman (ed) *General Purpose Technologies and Economic Growth*, MIT Press, Cambridge, Massachusetts (pp. 15-54).
- _____ (1998b), "The Consequences of Changes in GPTs" in E. Helpman (ed.) *General Purpose Technologies and Economic Growth*, MIT Press, Cambridge, Massachusetts (pp. 193-218).
- Lucas, R. E. (1988) "On the mechanics of economic development", *Journal of Monetary Economics*, 22, 3-42.
- Mairesse, J. and M. Sessenou (1991), "R&D and productivity: A survey of econometric studies at the firm level" *NBER Working Paper*, No. 3666.
- Nelson, R. and S. Winter (1982), *An Evolutionary Theory of Economic Change*, (Cambridge, Mass. : Belknap Press of Harvard University Press).
- Nye, D. (1990), *Electrifying America: Social meanings of a new technology, 1880-1940* (Cambridge and London: MIT Press).
- Pfaffenberger, R. C. and J. H. Patterson (1987), *Statistical Methods for Business and Economics* (Illinois: Irwin).
- Romer, D. (1996), *Advanced Macroeconomics*, (New York : McGraw-Hill Companies)
- Romer, P. (1986), "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, 94(5), 1002-37.
- Romer, P. (1990), "Endogenous Technological Change", *Journal of Political Economy*, 98(5), Part 2, S71-S102.

Rosenberg, N. 1982. *Inside the Black Box: Technology and Economics*. Cambridge: Cambridge University Press.

Rosenberg, N. (1994), "Why in America", in N. Rosenberg *Exploring the Black Box*, (Cambridge: Cambridge University Press), 80-108.

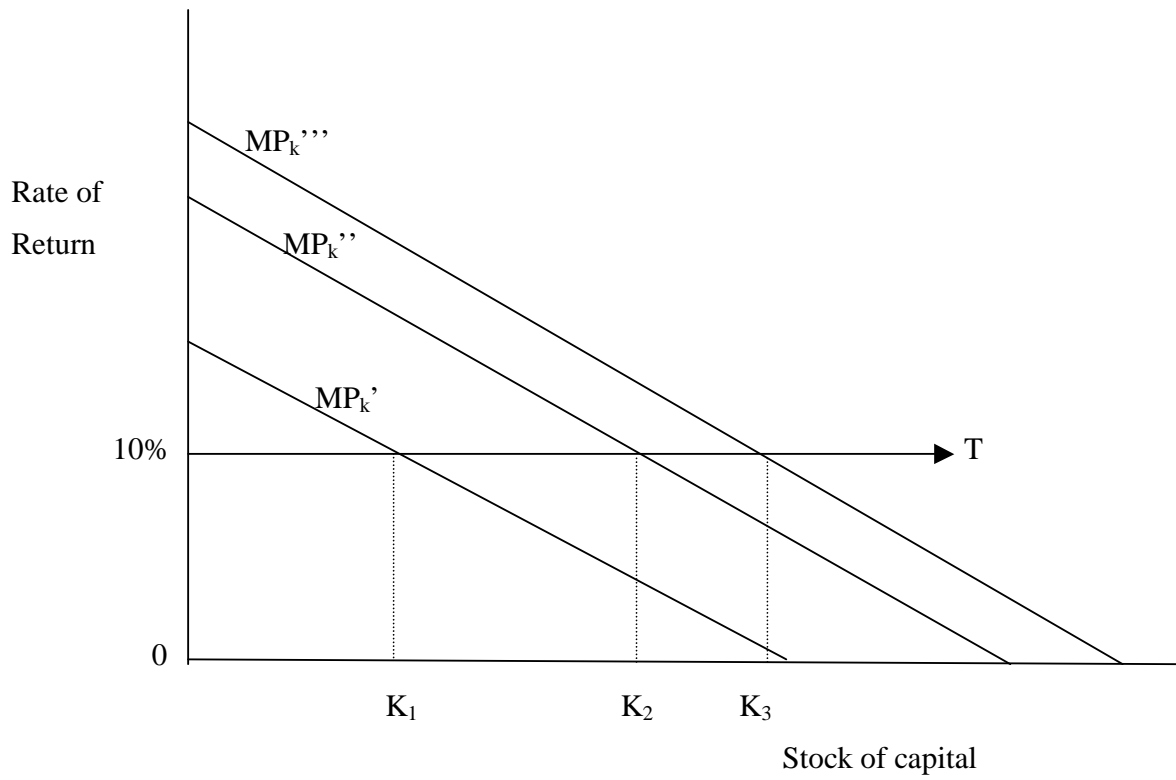
_____ (1997/63), "Technological Change in the Machine Tool Industry, 1840 – 1910", in *Economic Growth in the Long Run*, Vol. 3. 233-62 van-Ark (ed)

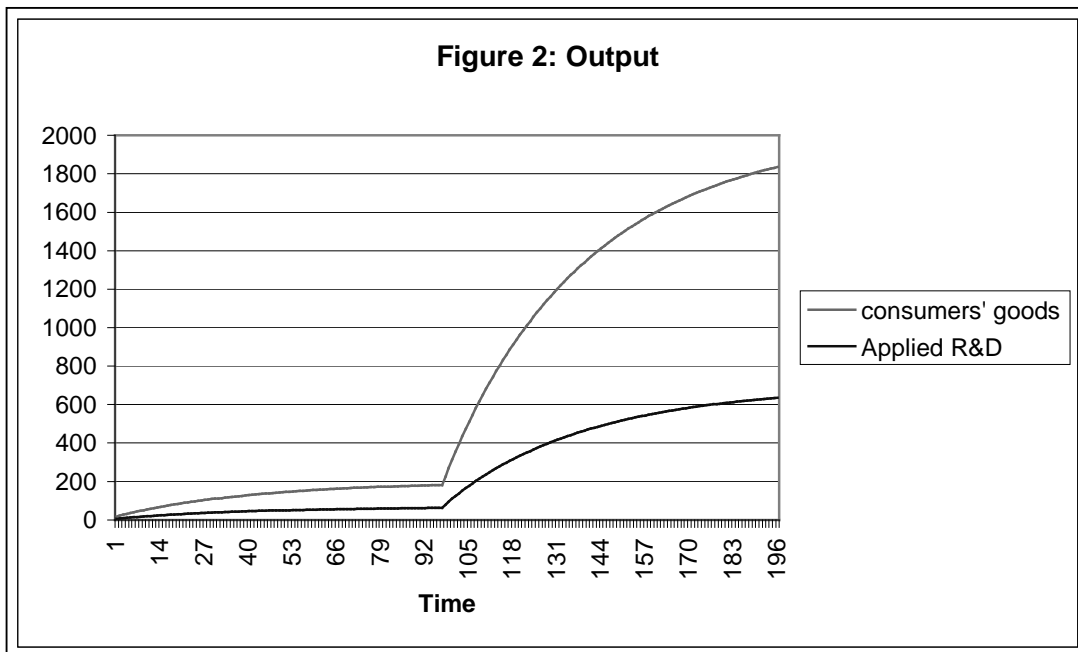
Schurr, S., *et al*, (1990), *Electricity in the American Economy* (New York: Greenwood Press).

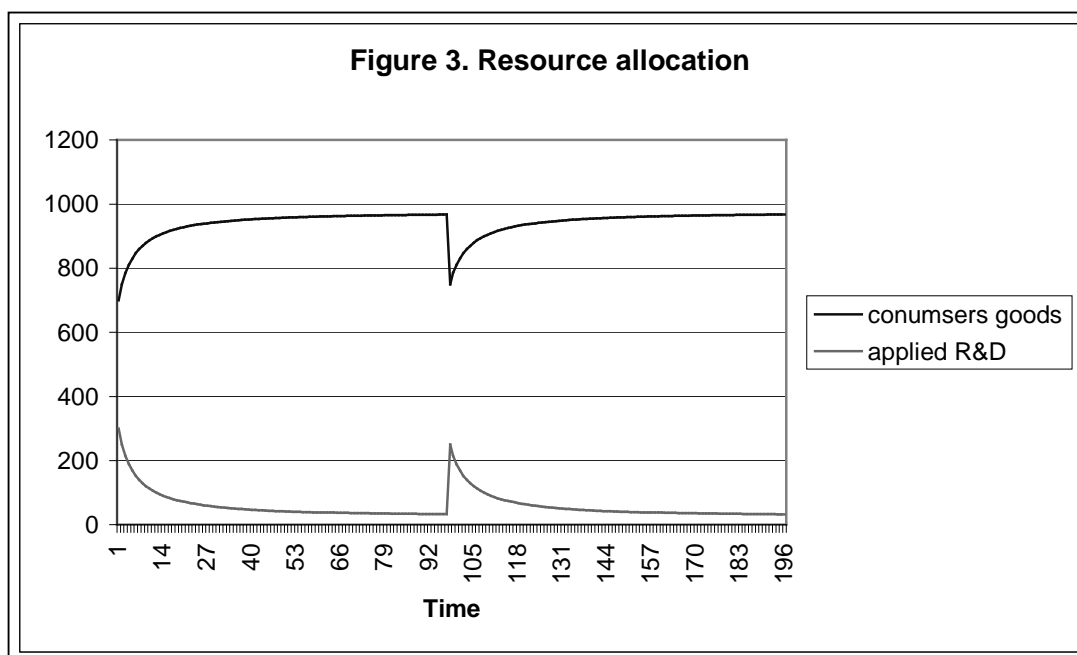
Schmookler, J. (1966), *Invention and Economic Growth*, (Cambridge: Harvard University Press).

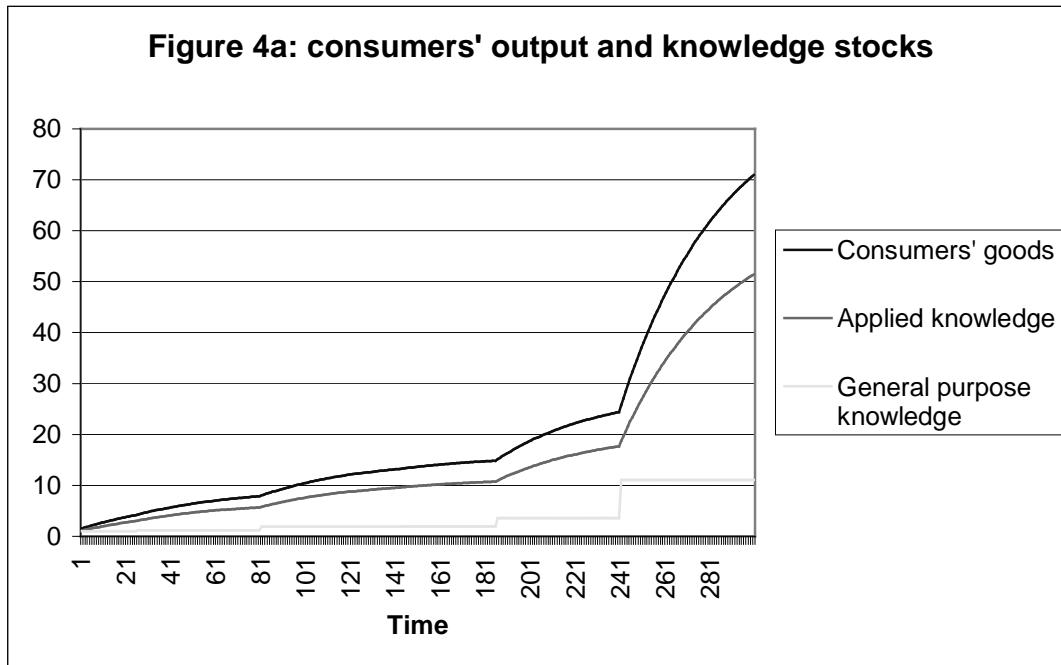
Womack, J.P., D.J. Jones and D. Roos (1990), *The Machine that Changed the World*, (New York: Rawson Associates).

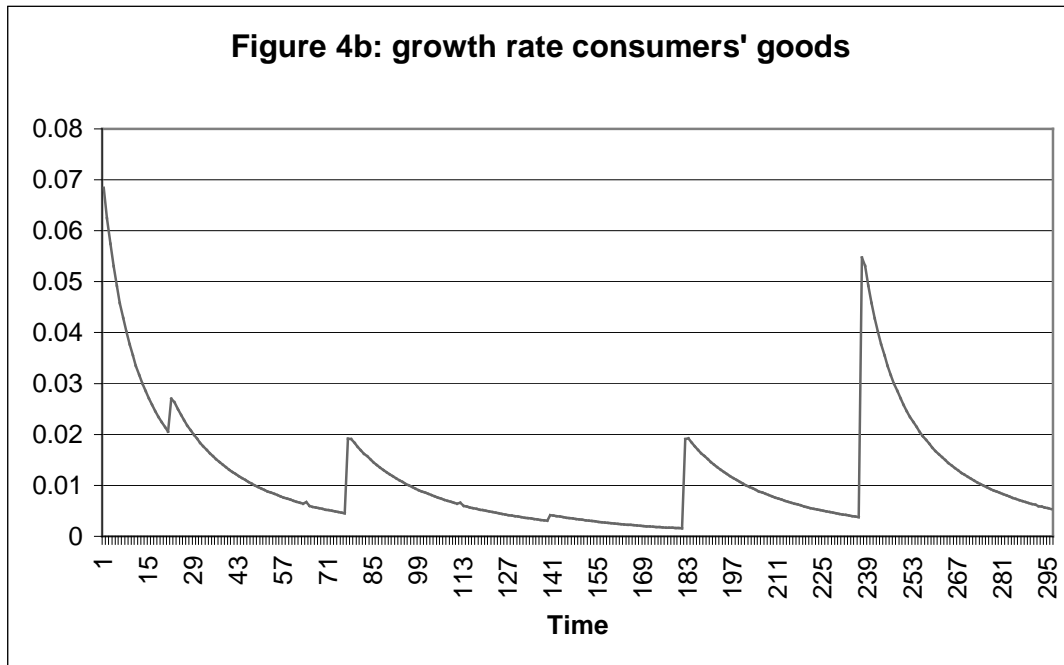
Figure 1

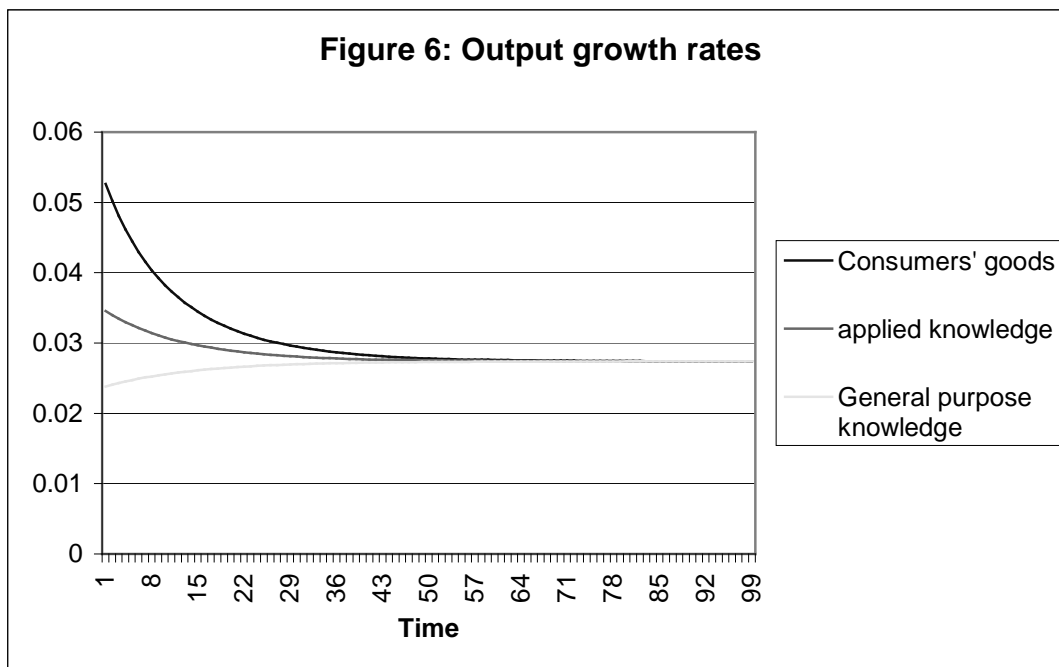
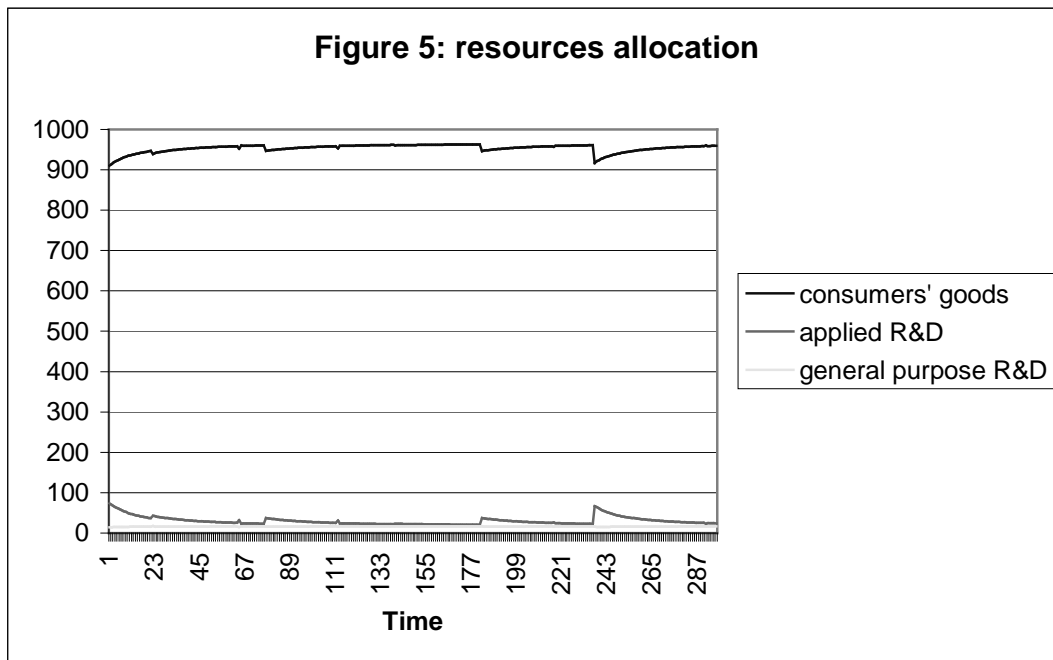


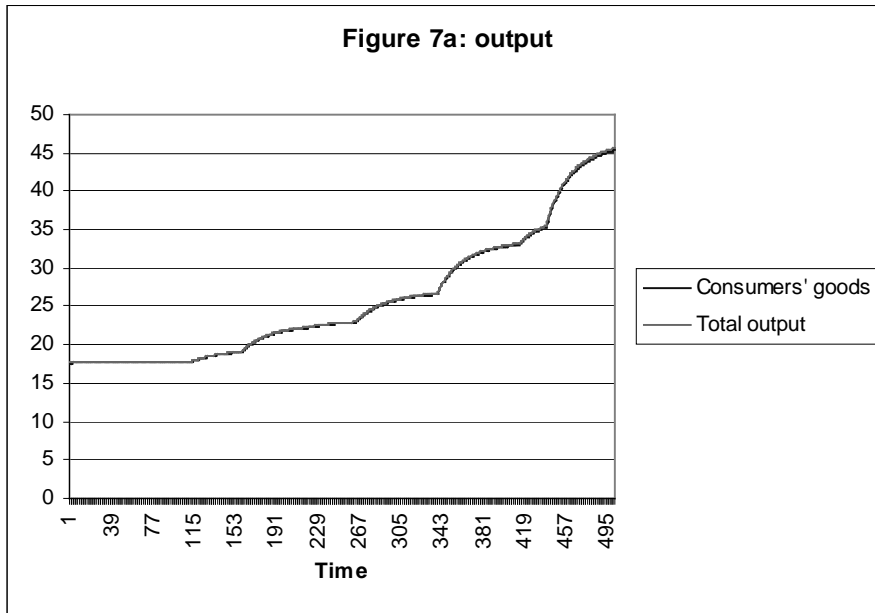


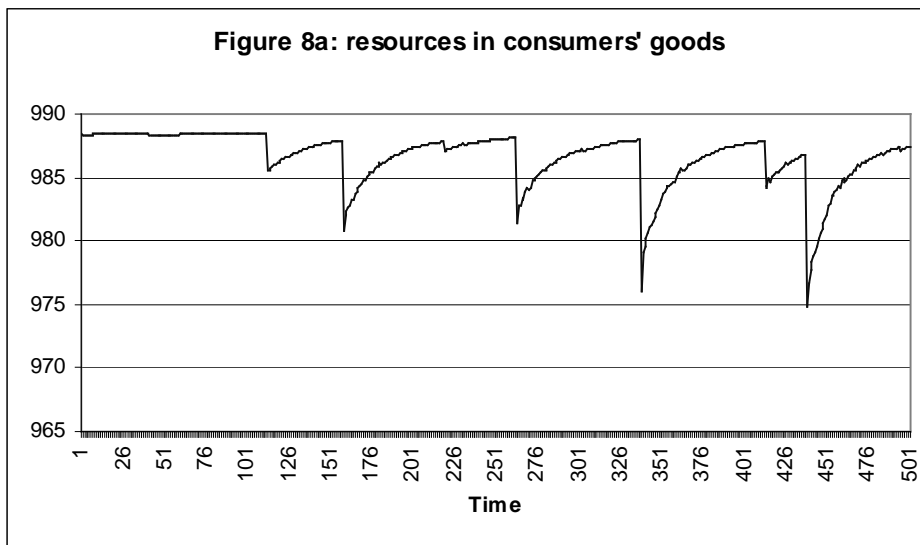
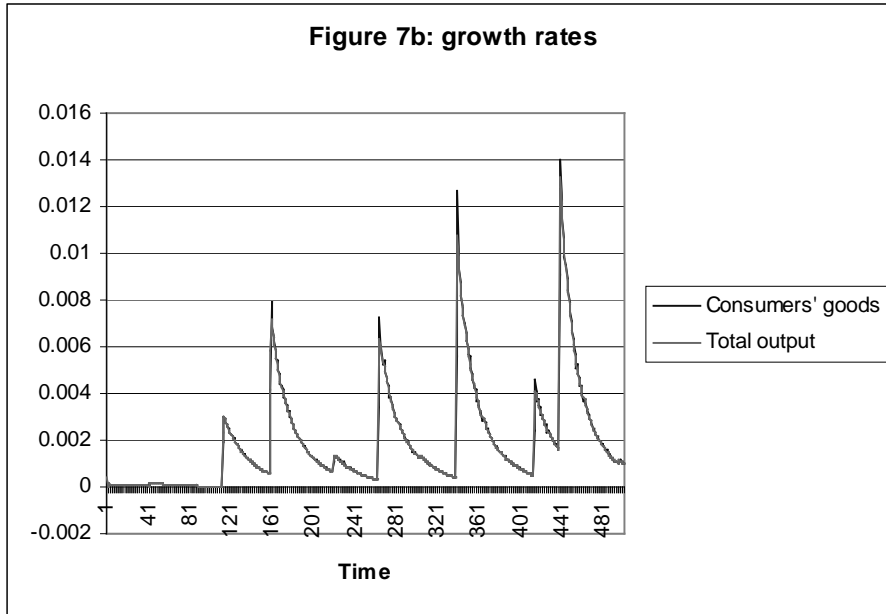


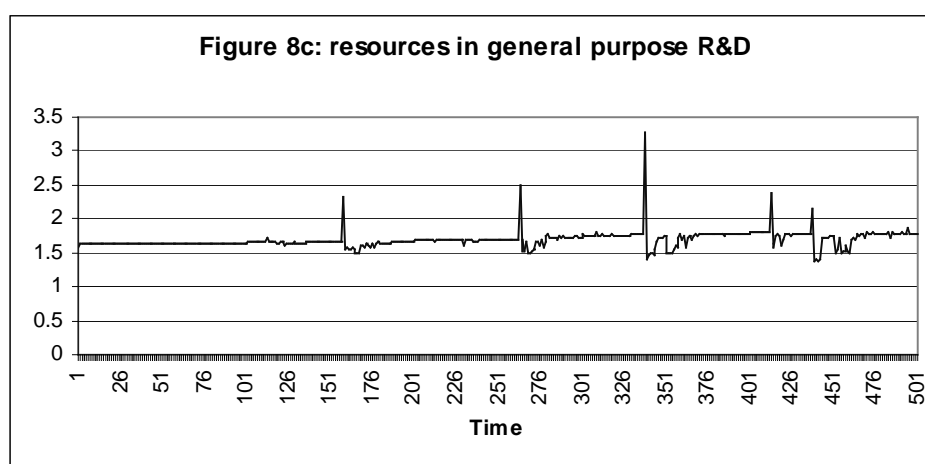
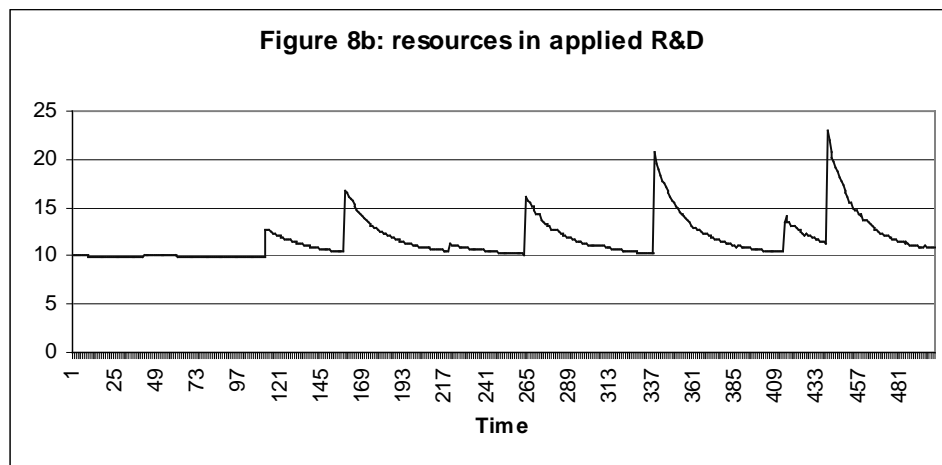












¹ For a detailed consideration of these characteristics and a development of the definition that follows in the text see Lipsey, Bekar and Carlaw (1998a).

² See, for example, Lipsey and Bekar (1995) and Lipsey, Bekar and Carlaw (1998a) for past technologies, Lipsey (2000) for some future ones, and Lipsey Bekar and Carlaw (1998b) for some generalisations about the impact of new GPTs.

³ In other cases, the quantities of one or more of the inputs used by *A* may appear in *B*'s production function, in others the *A*'s actions alter the prices of *B*'s inputs.

⁴ We are confining ourselves to positive externalities.

⁵ These pure cases of static and dynamic externalities are all that we need for present purposes. However, cross cases are also conceivable. First, an alteration by the initiating agents of some variable found in a standard GE model (a static case) could provide the opportunity for a technological change to be made by the receiving agents (a dynamic change). This common case occurs, for example, when increased consumption of some input increases its price sufficiently to provide the incentive for other agents to invent new substitutes. If the affected agents make returns in excess of opportunity costs from the innovation, they would presumably have been willing to pay something to induce the initiating agents to alter their consumption. Second, a change in the initiating agents' technology (a dynamic change) may lower the price of some inputs and so provide the opportunity for profitable substitution of inputs by the affected agents (a static case). Maximising affected agents would have been willing to pay something to the initiating agent to induce them to make the change in their technology.

⁶ If some agent has a patent on such a past invention, his actions with respect to that *patent* may upset optimum conditions but no action with respect to *the invention itself* can do so since its creation is a bygone.

⁷ Several technical problems to do with the firing of the engines and the running of wheels over tracks had to be solved first.

⁸ We use the terms invention and innovation interchangeably here because the distinction, although important in many contexts, is not important for the analysis of this paper.

⁹ Later in this part, we discuss the estimations that have been made by applying these procedures.

¹⁰ Mairesse and Sessenou (1991) put it this way: "Estimates based on firm data are sometimes considered to correspond to the private rate of return, whereas those obtained from aggregate data (at the industry level, for example) are seen as relating to the social rate of return. Another option here is to measure inside and outside R&D intensity and calculate two different rates of return. In that case the outside rate of return being positive is evidence of the externality."

¹¹ Aside from the conceptual problems that we have raised, this method requires that there be sufficient independent variation in inside and outside R&D to be able to separate their effects. For example, consider two firms. Let an increase in each firm's own R&D of \$1 increase its own output by \$0.10 and the other firm's output by \$0.05. Let the two firms be engaged in active competition and let their R&D's be fairly well synchronised. Now the elasticity of own output with respect to own R&D may pick up much of the external effect to give a measure close to \$0.15 for each \$1 of own R&D.

¹² Also, since the course of technological change over a century is largely unforeseeable being subject to genuine Knightian uncertainty, there is no real market failure even if, possessed with (the impossible) perfect foresight, planners would have chosen the statically non-optimal allocation of resources in order to obtain the superior dynamic path of technological change.

¹³ We realise that any major innovation is improved over time. Nonetheless, the benefit accrues in discrete modules rather than as a continuous flow as does say the benefit when, e.g., agent A's bees help to pollinate B's orchard.

¹⁴ This can be put in terms of the structuralist-evolutionary decomposition that we have used elsewhere. (See, for example, Lipsey Bekar and Carlaw 1998b.) We distinguish technological knowledge, the idea set for how to create economic value, from the facilitating structure, which embodies technological knowledge (among other things). Technological knowledge can be accumulating over long periods of time, for example, refining blueprints and introducing new and better components. However, it is the embodiment of these things in physical and human capital that occurs discretely and it is at this point that the receiving agents can see the new technology and that the spillover occurs. For example, the dynamo was the discrete embodiment of a great deal of evolving technological and scientific knowledge about electricity, while the turret lathe was the embodiment of a similar evolution of knowledge about machine tools.

¹⁵ Standard GE theory (constant tastes and technology) would predict that if some current agent obtained a patent on some long past invention all that would be affected would be the distribution of income, since the invention is a bygone. But if, say the wheel, were to become expensive to use, this would affect the course of current R&D and the path of future technological development. In a world in which technological change is endogenous, property rights over bygones do matter and can affect the course of future development.

¹⁶ Of course, this alteration in historical experience would have had untold further repercussions, that we have no hope of predicting. We engage in this thought experiment only to heighten the argument that these are not externalities in any operative sense.

¹⁷ This is in contrast to the Hicksian concepts of complementarity (and substitutability) in production theory that refer to the signs of the quantity responses of some item to a change in the price of some other item.

¹⁸ For a discussion of some of these spillovers see Rosenberg (1994) Chandler (1977), and for a discussion of the shortcomings of econometric estimates of the railway by counterfactual analysis see Freeman and Louca (2001, Chapter 1)

¹⁹ This example is based on Rosenberg (1994)

²⁰ Of course, it was never technically impossible for the Europeans to abandon their craft methods and innovate to the American system. But barring major restructuring in command-style economies, market economies, do not, for very good reasons, undergo such radical changes in their technology systems. Instead, they evolve incrementally along trajectories that are seldom discontinuous.

²¹ Was there a negative externality for Britain? Not necessarily. The choice of craft methods was a static maximising choice. Whether or not it would have paid to adopt US methods at higher cost in order to get a more favourable path of future technological change would depend on the social rate of discount.

²² This example is based on Womack, Jones and Roos (1990).

²³ Once parts manufacturers learned to produce standardised parts to a high degree of tolerance, in-house production of all parts proved inefficient. The system of large oligopolistic assemblers and many small competing parts manufacturers was then established first in the auto industry and then in most manufacturing industries where complex products were assembled from many parts.

²⁴ This example is mainly based on Nye (1990) and Schurr (1990).

²⁵ It took well over 200 years of cumulative research to complete the West's research agenda of understanding electricity and magnetism of which the following are just some of the highlights: (1) in 1670 Otto von Guericke invented a machine to produce an electric charge; (2) at the start of the 18th century Du Fay showed the difference between positive and negative electric charges; (3) the earliest form of condenser, the Leyden jar, was invented in 1745; (4) in 1752, Benjamin Franklin showed that atmospheric electricity was identical in form to the charge produced by a Leyden jar; (5) in 1766 Priestly proved that the force between electric charges varies inversely with the distance between the charges; (6) De Coulomb subsequently invented an instrument to measure electric charges accurately; (7) in 1800, Volta produced

the first electric battery; (8) in 1819, Oersted demonstrated that a magnetic field existed around an electric current; (9) in 1831, Faraday demonstrated that a current flowing through a coil of wire could induce a current in a nearby coil (he also developed the theory of electric lines of force); (10) in 1840, Joule and von Helmholtz demonstrated that electricity was a form of energy and that it obeyed the law of conservation of energy (Joule also showed that the magneto converts mechanical energy into electrical energy); (11) in 1845, Wheatston and Cooke patented an electro magnet to replace a permanent magnet in telegraphs; (12) in 1866 Wilde described a machine that used an electromagnet to turn unlimited amounts of mechanical energy into electrical energy; (13) in 1887, Wheatstone and Siemens invented a practical dynamo. The electric engine had arrived.

²⁶ This example is based on Rosenberg (1997/63)

²⁷ Of course, we have no doubt that there sometimes are enormous returns based on externalities. In spite of popular mythology, the great industrial fortunes of the late 19th and early 20th century were, as are the great ICT fortunes being made today, returns to getting something right in the development of new technologies. Sometimes, as with Ford, it is superior understanding of the power of new innovations; sometimes, as with Edison, it is inventive ability; sometimes, as with Gates, it is first mover advantage exploited to the utmost. Our point is that the beneficial dissemination of the full set of new technologies that stem from a GPT carry their own costs while the benefits are independent of any externalities or measured gains in TFP.

²⁸ We have actually produced a similar result using the simulation model presented in part three. In one version of the model we hold technology constant and allow agents to accumulate the whole set of inputs that go into production. These are subject to diminishing returns for all of the reasons we have just discussed. We calculate the rate of return through time and show that it falls throughout the simulation period. We then conduct the contrasting experiment holding resources constant and accumulating technological knowledge in the model with constant returns to accumulated knowledge in each sector, with the result that the rate of return is positive and constant. We generate an interesting result in the third version of our model that has a spillover for current GPTs to future GPTs in that the rate of return in the sector experiencing the spillover is growing through time, while the rate of return in the other two sectors is falling.

²⁹ We have discussed these issues in detail in the Introduction to Lipsey & Eaton (1997).

³⁰ Lipsey and Eaton (1989) discuss the lack of correspondence between observed facts and the Dixit-Stiglitz-Chamberlin version of monopolistic competition.

³¹ Lipsey (2001, forthcoming) discusses this kind of tolerated anomaly that pervades modern economics.

³² In subsequent models we introduce the accumulation of physical capital and add a fourth capital-producing sector.

³³ We are suppressing a complicated aggregation problem in the real world data by assuming that the scalar quantity in each of the right-hand side variables represents and aggregation over many sub-activities within each sector.

³⁴ In Lipsey and Carlaw (2001 forthcoming) we deal with the consequences of the usual assumption that resources such as land and raw materials, which cannot be accumulated, are included along with physical capital, which can be accumulated, in a composite variable called capital that is used in virtually all aggregate production functions. .

³⁵ An alternative but equivalent behavioural assumption is to view resources being allocated as if a social planner made the choices. If the maximisation is viewed as being undertaken by a social planner then the opportunity cost of using the resources is completely internalised and there are no externalities.

³⁶ Excel has a built-in worksheet routine called Solver that employs the Generalised Reduced Gradient (GRG2) non-linear optimisation code developed by Leon Lasdon, University of Texas at Austin, and Allan Waren, Cleveland State University. Portions of the Microsoft Excel Solver program code are copyright 1990, 1991, 1992, and 1995 by Frontline Systems, Inc. Portions are copyright 1989 by Optimal Methods,

Inc. We simply iterate through the maximisation problem above using Solver to maximise total output by allocating resources in each period.

³⁷ All of the figures are presented at the end of the paper.

³⁸ The simulation was actually run for 200 periods, with the first four iterations used to create initial values for the variables.

³⁹ It is necessary for modelling, and useful from a conceptual point of view, to lag the effect of the stock of applied R&D on the general purpose sector by a period. This is necessary in the model since the output of the general purpose sector determines the current stock of applied R&D. From a conceptual point of view changes in applied R&D will take time to have affect in the general purpose sector.

⁴⁰ For the moment the process of how this knowledge is rendered obsolete is treated as exogenous. A more sophisticated version of the model would make obsolescence endogenous and a function of the arrival of new GPTs.

⁴¹ Note the random variables λ , θ and ϑ are function of time, but we drop t subscripts in text to ease exposition.

⁴² With skewed distributions many small (close to zero) values are generated and occasional large values are generated.

⁴³ For $a = 1$ the mean of this skewed distribution is $E(z) = \int_0^1 zf(z)dz = \int_0^1 nz^n dz = \frac{n}{n+1}$. The median is 0.5.

⁴⁴ For $a > 1$ and $n > 1$ our procedure will actually produce a bimodal distribution. Thus a is restricted to be equal to one for all of the simulations.

⁴⁵ So far, utility has been implicit in the maximisation of consumption, which requires either that individual agents' utility, or social utility, be monotonically increasing in consumption. Applied and general purpose R&D can contribute to increases in consumers' goods but there is an opportunity cost in terms of consumption foregone today for more consumption tomorrow. This is the standard opportunity cost in all general equilibrium aggregate models.

⁴⁶ Many interesting variations arise by altering how this expectation is formed, e.g., from past evidence of the rate at which resources devoted to the GPT sector actually produce GPTs, or from subjective expectations of the success that each attempt will have.

⁴⁷ There are doubts about how well growth rates are measured given the changes in type of product, product quality, and the increasing servicisation of the economy. See Griliches (1984) on this point.

⁴⁸ Smookler (19qqq) and Rosneberg (1982) were some of the first to produce strong micro-economic evidence long before endogenous technical change entered macro growth models. Rosenberg's 1982 publication is a collection of key essays published over the preceding two decades

⁴⁹ Solow's original model produced growth in aggregate GDP and constant per capita GDP. By re-labelling Solow's growth in labour hours as growth in efficiency units of labour, a steady state growth path became one in which per capita GDP grew. This became the standard assumption of neoclassical growth models because of this result not because any empirical evidence suggested it was an empirical reality.

⁵⁰ Of course most endogenous growth models concerned with human capital or other forms of knowledge accumulation recognise that technology is not being held constant. The issue is whether it is appropriate to treat technological knowledge in the same manner as other inputs in production.

⁵¹ We have held these factors constant for the simple case but, at the cost of quite a bit of analytical complexity, we could have introduced optimal saving and capital accumulation. However, as we stated when we defined R as long as we imposed diminishing returns to all the factors of production, our results would be unchanged.

⁵² Our definition of a complementarity tells us that if “...arises in any situation in which the past or present decisions of the initiating agents with respect to their own technologies affect the value of the receiving agents’ existing technologies and/or their opportunities for making further technological changes.”