WHAT DOES TOTAL FACTOR PRODUCTIVITY MEASURE?

By
Richard G. Lipsey
Emeritus Professor Economics Simon Fraser University and
Fellow Canadian Institute for Advanced Research
and
Kenneth Carlaw
Lecturer, University of Canterbury, New Zealand
(k.carlaw@econ.canterbury.ac.nz)

Study Paper Version 02
January 18, 2001
REPLACES VERSION DATED 24 December 2000

Richard G. Lipsey
Simon Fraser University at Harbour Centre
515 West Hastings Street
Vancouver, BC
V6B 5K3
Voice: (604) 291 5036,
Fax: (604) 291 5034
email: rlipsey@sfu.ca,
home page: http://www.sfu.ca/~rlipsey

TABLE OF CONTENTS

I. INTRODUCTION......1

DIVERGENT VIEWS......1
PREVIEW......3
THE TECHNOLOGICAL BACKGROUND......4

II. THE BASICS OF TOTAL FACTOR PRODUCTIVITY......7

THE AGGREGATE PRODUCTION FUNCTION......7
DEFINITIONS......7
SOME WELL -KNOWN CONCERNS ABOUT TFP......9
Griliches......9
Aggregation of variables......10

III. CONCEPTUAL ISSUES......11

EMBODIED TECHNOLOGICAL CHANGE IN THE STANDARD EQUILIBRIUM FRAMEWORK......11
NATURAL RESOURCES MADE EXPLICIT......15
Background......15
Modelling resources......16
TIMING OF COST REDUCT IONS......18
An example......18
New products and processes more generally......19
Conclusion......20
A better measure......20
AGGREGATION OUT OF STATIC EQUILIBRIUM......20
DISTINGUISHING AND MEASURING TECHNOLOGICAL KNOWLEDGE AND HUMAN CAPITAL......24
Appropriate depreciation......24
Accumulation......24
R&D, the capital stock, and output......26
Human capital......27

IV. EXTERNALITIES, TECHNOLOGICAL COMPLEMENTARITIES AND EQUILIBRIUM......28

GROWTH THROUGH A SUCCESSION OF GPTS......28
DEFINITIONS......30
Externalities......30
Technological complementarities......33
The relation between complementarities and externalities......34
TECHNOLOGICAL COMPLEMENTARITIES NOT EXTERNALITIES DRIVE LONG TERM GROWTH......35
MEASUREMENT......35

V. A NON-EQUILIBRIUM MODEL OF SUSTAINED GROWTH......39

VI. CONCLUSIONS......42

APPENDIX A: TYPES OF TECHNOLOGICAL COMPLEMENTARITY......44

Type 1......44
Type 2......44
Type 3......45
Type 4......45
Type 5......46

FIGURES......48

BIBLIOGRAPHY......51

WHAT DOES TOTAL FACTOR PRODUCTIVITY MEASURE?

I. INTRODUCTION

We have two main concerns in this paper. First, we wish to assess the extent to which total factor productivity (TFP) can be taken as a measure of an economy's long-term technological change or technological dynamism. Second, we wish to better understand the complex set of technological interrelationships that we believe underlie the process of long-term growth. These go far beyond externalities as they are conventionally defined and measured.[1]

Divergent views

We do not believe that we are alone in being uncertain as to what TFP actually measures. The following quotations illustrate some of the different interpretations that various eminent economists have given to TFP measures. We list them in descending order of the scope that they give to TFP:

> (1) "A growth-accounting exercise [conducted by Alwyn Young.] produces the startling result that Singapore showed no technical progress at all." (Krugman: 55) "Singapore will only be able to sustain further growth by reorienting its policies from factor accumulation toward the considerably more subtle issue of technological change." (Young: 50)

These quotes assume that low TFP measures for Singapore indicate that during the period when its per capita income rose from third-world levels to those of industrialized countries, it underwent no technological change and was thus on an unsustainable path such as that followed by the USSR.

> (2) [T]otal factor productivity of an economy only increases if people 'work smarter' and learn to obtain more output from a given supply of inputs. Improvements in technology – the invention of the internal combustion engine, the introduction of electricity, of semiconductors – clearly increase total factor productivity." (Law: 6&7)

This quote says that TFP measures all improvements in technology, including such things as the introduction of electricity and the motorcar.

[1] This paper replaces the very preliminary version presented to the Statscan workshop 2 May 2000, which, as predicted in the introductory paragraph to that paper, did contain several errors.

> (3) "Technological progress *or* the growth of total factor productivity is estimated as a residual from *the* production function…. Total factor productivity is thus the best expression of the efficiency of economic production and the prospects for longer term increases in output. (Statscan 13-568: 50-51, italics added)

This quote tells us that TFP measures the effects of technological change and increases in efficiency over long periods of time.

> (4) "The central organising concept…[is] the division of observed growth in output per worker into two independent and additive elements: capital–labour substitution, reflected in movements around the production function; and increased efficiencies of resource use, as reflected by shifts in this function. To maintain additively, …the analysis…could not be applied cumulatively without introducing an interactio n term between capital substitution and increased efficiency....[The residual debate] never did attempt to answer the question, of what is the residual composed. This remains the dominant question (Metcalfe: 619-20)

In contrast, to the previous quote, this quote warns that TFP measures are only valid over relatively short periods of time and that there is some confusion about what TFP actually measures.

> (5) "The defining characteristic of [total factor] productivity as a source of economic growth is that the incomes generated by higher productivity are external to the economic activities that generate growth. These benefits "spill over" to income recipients not involved in these activities, severing the connection between the creation of growth and the incomes that result." (Jorgenson, 1995 pp. xvii.)

In direct contrast to the first three quotations, this one tells us that TFP measures only externalities and other free gifts associated with economic growth.

> (6) "Is there something possibly wrong with the way we ask the productivity question, with the analytical framework into which we force the available data? I think so. I would focus on the treatment of disequilibria and the measurement of knowledge and other externalities." Griliches (1994) (emphasis added).

This quote expresses even stronger reservations than the previous one, saying that a static framework may provide a misleading way of looking at productivity changes because they typically arise from externalities associated with disequilibria.

> (7) "The pioneers of this subject were quite clear that this finding of large residuals was an embarrassment, at best 'a measure of our ignorance'". (Griliches, 1994: 1)

This quote goes even further, cautioning us that TFP is a measure of our ignorance; it is nothing more than a measure of what we do not know.

It is an understatement to say that all of these quotations cannot be correct; TFP clearly means different things to different informed observers. Surely it is something close to a scandal that a measurement that is so much relied on for so many purposes seems to be so poorly understood.

We discern three main positions in these quotations. We summarise these below, attaching to each the names from our quotations that seem, more or less, to be associated with it.

- One group holds that changes in TFP measure the rate of technical change. (Law, Statscan, Krugman, Young.) We refer to this as the "conventional view".
- The second group holds that TFP measures only the free lunches of technical change, which are mainly associated with externalities and scale effects. (Jorgenson, and Griliches (6)) We refer to their position as the "J&G view".
- The third group is sceptical that TFP measures anything useful. (Metcalf, and Griliches (8))[2]

Our position is close to the J&G view. We disagree with them, however, on a number of smaller matters described in Section III. Section IV gives our more important disagreement with the J&G position. Here we argue that the really important external effects that are created by technological change are measured by neither the conventional definitions of externalities nor of TFP.

Preview

We are interested in long run technological change. We have argued for many years that much of the world has been living through a series of massive technological shocks that are associated with what is called information and communications technologies (ICTs).[3] Some economists doubt this judgement because, among other things, there was no evidence of big technological shocks in the measurements of total factor productivity (TFP) over the 1990s. If anything, judging from the TFP figures, technological change seems to have slowed over the last two decades of the 20th century.[4]

[2] Although we have only two representatives of this view in our quotations, it has many well-known members including the Cambridge economists who attacked the validity of the concept of an aggregate production function.

[3] Early statements of such a view are in Lipsey (1992), Lipsey (1993) and Lipsey and Bekar (1995).

[4] In our writings, we have advanced three main reasons why low rates of growth in TFP do not signify low rates of technological change. First, there are long lags between changes in technology and changes in productivity (see e.g., Lipsey Bekar and Carlaw 1998a and 1988b). Second, there is no necessary relation between successive changes in production regimes and the productivity changes associated with each. (See Lipsey and Bekar 1995: 66-69.) Third, we do not believe that TFP measures the rate of technological change. In this paper, we only consider the third reason. To consider the other reasons, we must break away from the neoclassical growth model where technological change is only observable by its results in changing TFP and use a model where the technology and productivity changes are not definitionally related to each other (as they are in aggregate growth models). This we have tried to do in the other references given in this footnote.

In this paper, we consider this objection. We first introduce some basic concepts concerning technological change and ask what it means conceptually to hold technology constant while accumulating capital. In section II, we outline the concept of total factor productivity and mention some well-known concerns relating to it. In section III, we express some of our own concerns related to embodied technical change, the treatment of natural resources, the timing of cost reductions, aggregating out of static equilibrium, and problem of measuring technical knowledge and human capital. In Section IV, we argue that the really important generalized benefits that are conferred by technological change are measured neither by conventional definitions of externalities, nor by TFP. There are, nonetheless, large and persistent benefits of technical change—benefits that we call technological complementarities. We also argue that these complementarities cannot be analysed in an equilibrium model of the sort that underlies most TFP measurements. This is a position with which Griliches seems to have some sympathy as shown by quote (6) above. Section V formalises a simplified version of the arguments given in section IV in a model in which endogenous technological change produces continuous growth in the *absence of* externalities, increasing returns, constant returns to the accumulating factors, and a balanced growth path. In Section VI we conclude that whatever may or may not be measured by TFP, it *cannot* be understood as measuring technological change or technological dynamism.

The technological background

We conceive of society as starting with two basic factors, labour, L, and the endowments provided by nature, R.[5] Society then produces output, Y, using these endowments along with two main bodies of created assets, physical capital, K, and human capital, H. The measurement of output poses massive problems, which we assume to be solved because we wish to focus on the input side.[6]

We use a wider definition of technological knowledge than is usual. For us, technological knowledge is the *idea set* of how to create economic value. This knowledge concerns product, process, and organisational technologies .What we call the facilitating structure is the *embodiment* of these ideas in such things as physical capital, human capital and organizational forms, all of which are all part that structure.[7]

[5] What is provided by nature is exogenous but similar resources may also be created by human effort as when a forest is replanted or a lake restocked.

[6] As we later argue, technological change prevents the decline of the marginal productivity of capital that would inevitably occur with constant technology. A similar force operates on the output side. If the technology of consumers goods and services had been held constant at those existing at some earlier time, say 1900, diminishing utility of income would be a reality, as consumers wondered what to do with a third and fourth horse and buggy and train trip to the nearby seaside. Technological changes in consumer's goods constantly present consumers with new consumption possibilities and remove at least the inevitability of declining marginal utility of income as income rises over time, (which does not prevent marginal utility of income fromdeclining at a point in time when, of necessity, technology is constant).

[7] In other writings, we have detailed this model and tried to show how it helps to understand why the speed of technological change need not be closely related to the rate of productivity growth. See, e.g., Lipsey and Bekar (1995) and Lipsey Bekar and Carlaw 1998b.

Economic historians and students of technology are agreed that technological change is *the* major determinant of long-term economic growth over the centuries.[8] So the problem of explaining growth over time and across countries is mainly one of explaining the generation, adaptation within one country, and international diffusion of, new product, process and organisational technologies. In the long term, these new technologies transform our standards of living, our economic, social and political ways of life, and even our value systems.

Much new technological knowledge is embodied in capital equipment whose accumulation is measured as gross investment. So technological change and investment are interrelated, the latter being the vehicle by which the former enters the production process. It follows that the long-term rate of growth will be slowed by anything that slows either the development of new technologies or the rate at which new technologies are embodied through investment (such as unnecessarily high interest rates). Thus the high correlation between new investment and growth does not demonstrate that investment is the main cause of long-term growth. Both technological change and investment are necessary.[9]

Since we are considering attempts to measure technological change, the concept of constant technology is critical. To hold technology conceptually constant while capital is accumulating, we need to do the following: hold all product, process, and organisational technologies constant at what was known at some base period; accumulate more physical capital that embodies the technologies then in use, or others that were known but not in use; accumulate more human capital in the form of more education in what was known at that time.

For a specific example, let the changes in human and physical capital be the measured accumulations that occurred between the base period of 1900 and the given period of 2000. Estimate the increase in output with constant 1900 technology. This is a measure of what could have been achieved by investment in physical and human capital without any alteration in technological knowledge. Now calculate the actual increase in output. The difference is “due to” or “enabled by” technological change in the sense that it could not have happened without such change. Measured over a long period of a century or more, the difference due to technological change would be very large indeed.

Here are just a few illustrative examples of what the constant-technology experiment would reveal.

[8] Other major sources are pure capital accumulation with constant technology and scale effects associated with increases in the size of markets.

[9] Nonetheless, consider the choice between two polar cases: One could either live in a society in which technology advanced but was only embodied through “replacement investment” since net investment (and hence measured capital accumulation) was zero, or in a society in which nothing was known that was not known in 1900 and more and more investment had been made in 1900-style productive facilities to produce 1900-style goods and services. We wager that most people would prefer the former alternative. As Solow long ago observed: “One could imagine this [growth] process taking place without net capital formation as old-fashioned capital goods are replaced by the latest models, so that the capital-labour ratio need not change systematically.” (Solow, 1957:316)

- Feeding 6 billion people with the agricultural technologies of 1900 would have been literally impossible.[10] Sooner or later, Malthusian checks would have become a reality, as ever expanding populations would have encountered increasing food shortages. (Among other things this shows that population and the labour force cannot be taken as independent of technology.)
- Pollution would have become a massive problem. By our standards, 1900 technologies were heavily polluting and to increase production sufficiently to employ all the new capital would have led to major increases in pollution.
- Exhaustion of specific resources would have become a serious problem. Most new technologies are absolutely saving in resources.[11] Thus, to produce the value of today's manufacturing and service output with 1900 technologies would have required vastly more resources than are currently used. Further, with no changes in technological knowledge the scope for replacement of scarce resources by plentiful ones would have been greatly restricted.

In another demonstration of these points, the calculations of the Club of Rome in the 1970s showed the folly of believing that production could long be increased at current rates with no change in technology. The Club's predictions of doom were falsified by continued technological advance, which invalidated their calculations on resource exhaustion and unsustainable pollution. But these mistaken predictions do show for how few decades current world growth rates could be sustained in a world of static technology.[12]

[10] Of course, population is endogenous and it is not clear how much population would have increased if food-producing technologies had remained frozen at their 1900 levels. However, Western sanitation, health and medical practices had already lowered death rates in the West and had led to large increases in life expectancy with a resulting population boom, and these practices were already being extended to the less developed countries. Thus, some large population expansion would certainly have occurred.

[11] This is a process that Grubler (1998) calls "dematerialization". Among the many illustrations that he quotes are these: during the period 1975-94 "Total materials requirements per unit of (constant) GDP have declined between 1.3% per year in Germany, 2% per year in Japan, and 2% per year in the Netherlands" (Grubler: 240).

[12] A further problem arises in altering the capital labour ratio with fixed technology. In so far as the population increase and new capital is merely replicating existing productive facilities staffed by new workers, manufacturing and service production can be expanded at more or less constant returns to scale. But this process employs more persons while leaving constant their productivity and hence their real wages and living standards. Raising living standards with static technology requires increasing the capital labour ratio. Although new technologies often do this, existing technologies, especially in manufacturing, typically have little scope for varying the capital labour ratio, factor proportions being more or less built into them. It would be impossible, for example, to take a plant designed in 1900 to produce steam engines and increase the amount of capital per worker by 500%. There was room for some substitution of capital for labour within the confines of existing technology of steam engine construction, but not much. So, if the economy were to grow by increasing the capital per head from its level in 1900 to its level in 2000 without altering technology, it would become increasingly difficult to find places in which the extra capital could be profitably employed. Much of it would end up in non-manufacturing activities, while manufacturing areas, which were, in actuality, some of the main sources of rising living standards throughout the 20th century, would be carried on in unchanged ways with little increase in capital per head.

II. THE BASICS OF TOTAL FACTOR PRODUCTIVITY

The aggregate production function

Both neoclassical and endogenous growth models employ an aggregate production function, a specific example of which is shown in equation II.1 below. Any such economy-wide function is a theoretical construct with no direct empirical counterpart in actual micro data. Conceptually, it can be aggregated from a perfectly competitive world, but not from the mixture of monopoly, oligopoly, monopolistic and perfect competition that characterises real-world industrial structures. Furthermore, the aggregation requires the existence of the *end state* of perfect competition rather than existence of the *process* of competition that we actually see in the real world.

This last point is important and requires elaboration. Standard neoclassical theory treats competition as an *end state* that is perfectly competitive equilibrium. The Austrian tradition 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 TFP measurements, and all other theories that use the neoclassical aggregate production function, assume the existence of end-state competition; 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 where the rigorous conditions cannot be fulfilled (which in aggregating production functions is more or less everywhere)—the judgement being that deviations caused by non-fulfillment are quantitatively unimportant.[13]

To make contact with the existing literature, we will proceed as if the aggregate production function is a meaningful concept, although this must be a matter of judgement not formal proof given the reality of process rather than end state competition.

Definitions

Consider the simple Cobb-Douglas version of the aggregate function:

[13] Lipsey (2001) discusses this kind of tolerated anomaly that pervades modern economics.

(II.1) $Y = AL^{\alpha}K^{\beta}\,,\ \ \alpha + \beta = 1$

Total aggregate output is measured as *Y. L* is an index of aggregate labour inputs. *K* is an index of aggregate capital. Typically *Y, L* and *K* are independently measured while *A*, α and β are statistical estimations. *A* is an index of the aggregate state of technology called total factor productivity. Since *A* is not a pure number, it carries no interesting information in itself. But changes in the number indicate shifts in the relation between measured aggregate inputs and outputs and *in this aggregate model* these changes are assumed to be caused by changes in technology (or changes in efficiency and/or in the scale of operations of firms).

The geometric index version of TFP is calculated by dividing both sides of the production function by:

$L^{\alpha}K^{\beta}\,,$

to produce a measure of TFP:

$$TFP = A = \frac{Y}{L^{\alpha}K^{\beta}}\,.$$

The growth rate measure of TFP is then calculated as an arithmetic index generated by taking time derivatives of both sides of the TFP expression:

(II.2) $\dfrac{\dot{A}}{A} = \dfrac{\dot{Y}}{Y} - \alpha\dfrac{\dot{L}}{L} - \beta\dfrac{\dot{K}}{K}\,.$

The dot superscript denotes the time derivative. *a* and *b* are the shares of output/income accruing to labour and capital. That is:

$\alpha = \dfrac{wL}{Y}$ and

$\beta = \dfrac{\pi}{Y} = \dfrac{rK}{Y}\,.$

where *w* is wages paid to labour, ***p*** is total profits and *r* is the real rental rate of capital. These shares imply that

$$\frac{wL}{Y} + \frac{\pi}{Y} = 1 \ \text{ or } \ \frac{wL}{Y} + \frac{rK}{Y} = 1\,.$$

If we have measures of the physical inputs of labour and capital (II.2) defines a Divisia index of inputs, which is the percentage change in each input weighted by its relative share in input costs.

Most work on TFP uses a Tornquist index, which is slightly different but is basically a percentage change index that averages base and given years weighted indexes as does the Fisher Ideal index. For our purposes, we use the simpler Divisia index, which weights percentage changes in specific inputs by their share of total cost. None of the conclusions we reach would be seriously affected by the substitution of one index for the other.

Some well-known concerns about TFP[14]

Griliches

Griliches (Palgrave: 1010-13) outlines some conceptual and empirical problems concerning the measurement of TFP. These relate to the following issues: (1) a relevant concept of capital, (2) measurement of output, (3) measurement of inputs, (4) the place of R&D and public infrastructure, (5) missing or inappropriate data, (6) weights for indices. (7) theoretical specifications of relations between inputs, technology and aggregate production functions, (8) aggregation over heterogeneity. Concerning point (6), Diewert (1987: 767-780) shows that very restrictive assumptions have to be satisfied to generate these indices of output and input,

Griliches rewrites the TFP expression, accounting for these potential errors, to argue that the measured number is so riddled with flaws as to be seriously misleading.

$$\text{TFP}=\frac{\dot{A}}{A}=s\left(\frac{\dot{K}^*}{K}-\frac{\dot{K}}{K}\right)+(1-s)\left(\frac{\dot{L}^*}{L}-\frac{\dot{L}}{L}\right)+(s*-s)\left(\frac{\dot{K}^*}{K}-\frac{\dot{L}^*}{L}\right)$$

$$+\,h\left[s*\frac{\dot{K}^*}{K}+(1-s*)\frac{\dot{L}^*}{L}-f\right]+\alpha_z z+\mu+t$$

The symbols marked with asterisks denote “correctly” measured inputs.

s is share weight of capital in total output and (1 - *s*) is the share weight for labour. These are functions of elasticities of output with respect to specific inputs.

$$s*=\frac{\alpha_k}{\alpha_k+\alpha_l}=\frac{\alpha_k}{1+h}$$

is the correctly measured share of capital. The α’s are the true elasticities with respect to the inputs. $h=\alpha_k+\alpha_l-1$ is a measure of the economies of scale with respect to the measured percentage rates of change of the conventional inputs (K and L). *f* is the rate of growth of establishments. *z* is the rate of growth of inputs which affect output, *but which are not included*. m is errors in measurement. Finally, *t* is the “true” rate of growth of the average level of disembodied technology, which includes externalities from technological change.

The first term:

$$s\left(\frac{\dot{K}^*}{K}-\frac{\dot{K}}{K}\right)$$

reflects the rate of growth in measurement error of capital.

The second term:

[14] Others have expressed concerns about what TFP does and does not measure. These concerns are mainly additional to those expressed in this article.

$$(1-s)\left(\frac{\dot{L}^{*}}{L}-\frac{\dot{L}}{L}\right)$$

reflects the rate of growth in measurement error of labour.

The third term:

$$(s*-s)\left(\frac{\dot{K}^{*}}{K}-\frac{\dot{L}^{*}}{L}\right)$$

reflects the error in assessing the relative contribution of each factor.

The fourth term:

$$h[s*\frac{\dot{K}^{*}}{K}+(1-s*)\frac{\dot{L}^{*}}{L}-f]$$

is the economies of scale term. It is zero if either *h* = 0, constant returns to scale, or the rate of growth in the number of new establishments (*f*) just equals in the total weighted input, which implies that growth of output is by replication of identical establishments[15].

The fifth term, $\alpha_z z$, reflects the contribution of omitted inputs (private or public).

The sixth term, μ, is unspecified errors.

The seventh term, *t*, is the pure residual term (i.e., the amount of growth not accounted for by the expanded list of possible sources).

Since we can be sure that there are errors in all of these terms, some of which may be quite large, we must be cautious in relying on any interpretation of changes in measured TFP. In a paper on “The Discovery of the Residual” Griliches concludes that “ [a]ll of the pioneers of this subject were quite clear about the tenuousness of such calculations and that it may be misleading to identify the results as ‘pure’ measures of technical progress.” (Griliches 1995: 6)

Aggregation of variables

The Divisia and Tornquist indexes aggregate by weighting percentage changes in quantities by expenditure weights. But the quantities used must almost invariably themselves be aggregates. When TFP is calculated from a macro production function, the “quantities” used are the aggregate capital stock and the aggregate labour supply; when it is calculated from industry data, they will be industry capital and industry labour; similarly for firms, it will be each firm’s capital stock and its employed labour. To get to basic quantities without any prior aggregation, extremely detailed micro data would be needed with a separate quantity input for each capital service, of which there are thousands in a typical manufacturing firm.

[15] Griliches refers to “*f*” as both establishments and firms at various parts of his argument. It seems to us that establishments, or plants, is the right concept. It is quite possible to replicate plants with no change in the number of firms so that his term would be zero when *f* is establishments and positive when *f* is firms.

Thus, no matter how disaggregated are the physical quantities that are used for any calculation of a TFP index, they are typically aggregated over some group of heterogeneous capital goods (or capital services) by converting them to values. To do this, prices or marginal productivities (equal in competitive end-state conditions) are typically used. (What else could be used?) This creates problems as illustrated by the following simple example.

Rewrite (II.1) to add efficiency parameters to *L* and *K*:

(II.3) $Y = A(mK)^{\alpha}(nL)^{1-\alpha}$.

Thus:

$Y_K = \alpha A m^{\alpha} K^{\alpha-1}(nL)^{1-\alpha}$

and

$Y_L = (1-\alpha)A(mK)^{\alpha} n^{1-\alpha} L^{-\alpha}$

(Subscripts indicate partial derivatives.)

Allowing only *m* to change:

$$\frac{\dot{Y}_k}{Y_k} = \alpha\frac{\dot{m}}{m} = \frac{\dot{Y}_L}{Y_L} = \frac{\dot{Y}}{Y}$$

So income rises at the same rate as measured capital inputs leaving TFP unchanged.

This is obvious and well known. If the increase in the value of the capital stock represents a return that just covers the R&D costs of the innovation that raised the efficiency of capital, then zero TFP is the correct answer on the "J&G view." If it is a free lunch, possibly coming out of the blue in an unexpected invention made at negligible cost, then the zero TFP is not the correct answer, even on the J&G view. The point is that to measure inputs by the value of their marginal product is to force TFP close to zero whatever the cause of the increase in marginal product. If we wish to test the extent of free lunches, we need to define TFP so that its value can be high or low depending on the strength of the free lunch effect.

III. CONCEPTUAL ISSUES

In this section, we consider a number of conceptual issues that suggest some use of extreme caution when interpreting TFP numbers as measures of technological change.

Embodied technological change in the standard equilibrium framework

It is well known that much technological change is concentrated in the industries producing capital goods. An excellent case study of such changes is found in Rosenberg's analysis of the US machine tool industry in the 19th and early 20th century. The development of cost-cutting, output-increasing tools of standardised production was one of the major reasons why the US began to overhaul Europe in economic growth and technological dynamism (Rosenberg 1976 and 1994).

Over the years, many economists have been concerned about whether or not the significant amount of technical change that is embodied in capital goods will show up in TFP numbers. Specifically, will technological change that is embodied in capital goods lead to an overstatement of the amount of change in capital and hence cause TFP measurements to understate the amount of technical change?

To study this issue, we assume two industries. The goods industry produces a constant amount of some final good, *Y*, whose specifications are unchanged. Its inputs are capital, which takes the form of machines, and labour. It has the following Cobb Douglas production function:

(III.1) $Y = A(mK)^{\alpha}(nL)^{1-\alpha}$

where K is capital measured in value of machines and L is labour measured in physical units, while *m* and *n* are efficiency parameters attached to capital and labour respectively.

The machines are produced by a capital goods industry. The producers spend money on R&D to create new technologies that alter the nature of the machines they produce. We follow Jorgenson in assuming that the new machine is sold by the capital goods industry at a price just sufficient to recover its direct costs of production and the R&D costs of creating it (the amount spent on R&D plus a return on R&D investment equal to the going rate of return on capital investment). In other words, the price of the improved capital good capitalises the cost of the R&D that created it. Thus the capital good producers will register no change in their TFP since their costs change in the same proportion as the price of their output.

We consider four different types of technological change. The first is a disembodied change in the organisation of production in the final goods industry that, with output constant, lowers both inputs by x%. The next three are embodied changes that originate in the capital goods industry and that alter the nature of the machines they produce. In each case, we chose our numbers so that the measured amount of capital increases by *x*%. In each case, we calculate TFP for the final goods producers. Notice that although it is common in the literature to start with certain parameter changes and identify them with certain types of technological change, we do the opposite. We start with certain types of real world technological change and identify them with the parameter changes that will express them.

Case 1 a disembodied technical change in the final goods industry lowers unit labour and capital costs in the same proportion: This kind of change occurs when the final goods producers reorganize production to produce the same output with a lower amount of inputs. The reorganization of the factory after the replacement of steam power by electric engines on each machine tool (the unit drive) is an example. If both inputs change by the same proportionate amount, *x*%, the change shows up as a rise in the efficiency parameter *A* in III.1 above by *x*%.

Assume, just to benchmark, that the physical volume of output is unchanged and consider two polar cases. First, let the organisational change be costless. It is an isolated stroke of genius on the part of management that involves no resource cost to conceive and institute. Now the value of output is unchanged while costs fall by *x*%. The industry's TFP will rise by *x*% Second, let the R&D cost of developing the industry's organisational change be the equivalent of x% of the resource cost of production, i.e., there is an x% additional input cost. Now overall costs, direct production and R&D, will be unchanged as will the value of output. The industry's TFP will be unchanged. Thus it is not true that disembodied technical change necessarily raises TFP. What matters is whether the fall in direct costs per dollar's worth of output is more, or less than, or equal to the cost of the R&D that developed the disembodied change.

Case 2 a new embodied technology saves on both labour and capital costs: This is the common case in which the capital goods industry develops a new technology that is absolutely saving on both labour and capital. Lean production (or Toyotaism) in the Japanese automobile industry is one of many examples. (Womack et al 1990) To be specific, we assume that the cost of producing the new machine is 2*x*% less that the cost of producing the old one. The R&D costs are just covered when the new machine is sold for *x*% less than the old machine. The new machine is also assumed to use *x*% less labour to produce the same amount of output as the old machine. This common type of innovation shows up in III.1 above as an increase in both of the efficiency parameters, *m* and *n*, by *x* percent. Thus with constant output, inputs of capital and labour (measured in physical units) fall by x% so that the industry's TFP rises by *x*%.

Notice that in equation III.1, as in any CRS production function, this change is indistinguishable from an increase in the parameter *A* by *x*%. So this typ e of embodied technological change looks analytically like disembodied change. Note also that mixed cases in which one efficiency parameter rises by *x*% and the other by *y*% and $x > y$, are indistinguishable from an increase in *A* by *y*% and an increase of $(x - y)$% in the efficiency parameter with the larger increase.

Case 3 the new embodied technology saves only on labour inputs : The new machine is assumed to produce *x*% more output using the same amount of labour, a type of invention that is common in the history of technological change. The R&D cost of the new machine can just be recovered by the capital goods industry if the machine sells for *x*% more than the old machine. When the consumers goods industry replaces the old machines with new ones, this uses savings equal to the additional value of the machines. So, as the accounting is normally done, the industry's capital stock will grow by *x%.* The industry will then have *x*% more measured capital and *x*% more output—and, therefore, an increase in the efficiency of labour as if there were *x*% more labour (i.e., the labour efficiency parameter, *n*, increases by *x*%). The result is a change in measured TFP: ΔTFP $= x\% - \alpha x\% = (1-\alpha)x\%$. This is Harrod-neutral technical progress with measured capital and efficiency units of labour both growing at *x*%. Note that the investment in better machines is real investment in the sense that if the resources had not been devoted to R&D, they could have produced an equal increase in output by being devoted to consumption goods.[16]

[16] It has been argued to us that the machinery producers could raise the price to the new machine so as to appropriate all of the extra gain for itself, leaving no change in TFP for the final goods producers. However, if TFP is measured sufficiently finely, all this does is to transfer the technological gain to the capital goods industries who get an extra increase in the value of their output with no offsetting increase in costs. The general point is that if there is a surplus of increased value of output over increased costs there is a TFP gain and all that various pricing policies do is to determine where in the economy it is located.

Case 4 the embodied technical change saves only on capital inputs: Now assume that the R&D produces new machines that act as if there were *x*% more of them than the old machines. Machines increase in efficiency but labour does not. Although this is a natural change to consider in theoretical models, it is hard to identify real world technological changes that have produced this result. The efficiency parameter on machines grows by *x*%, making the growth of the industry's output equal to αx%. The marginal product of machines falls steadily but the absolute number increases just enough to balance that, making *Y* grow at the constant rate αx%. TFP = αx -αx% = zero%.

Case 5 (no technical change): In cases 2-4, the measured amount of capital increases by *x*%. Now let capital increase by this amount with no technical change. (In this case, and in this case alone, the physical number of machines used must increase.). Analytically this is the same as Case 4. Measured capital increases by *x*%, making output and inputs increase by αx% giving no increase in measured TFP. It follows that the TFP measures in cases 1-4 show the margin of gain in allocating resources to R&D to produce newly designed capital rather than to produce more capital with unchanged specifications.

Conclusions: A number of conclusions are suggested by this analysis. Although some of these are well known to those who work in this field (but not necessarily to all those who use TFP figures), they all deserve emphasis in the context of this paper.

- It is incorrect to follow the advice sometimes given that R&D should be treated as a separate capital item and its stock accumulated as an element of capital in the national accounts. If the price of the capital goods rises to recover the R&D that created it, this capitalises the value of the R&D in the price of the product it helped to create. Treating accumulated R&D as a separate capital stock double counts it.

- In theoretical treatments, the distinction between embodied and disembodied technical change is often made in terms of the parameter of the production function that is affected, a shift in *A* being disembodied technical change, and a shift in either *m* or *n* embodied change. These shifts do not, however, map simply into the kinds of technological change we see in real world. We cannot in practice distinguish empirically between genuine disembodied technical change and technological change that is embodied in a new machine that lowers all input costs in equal proportion. Both appear in the observed function as Hicks neutral growth. The observation of Hicks neutral growth does not, therefore, imply that technological change is not embodied in new capital equipment. Furthermore, where it does occur, disembodied technological change does not necessarily provide the kind of free gift that increases TFP. The effect that it does have on TFP has nothing to do with being disembodied or embodied; all that matters is whether or not the gains more than cover the costs of creating them.

- The TFP measure in each of the four cases correctly reflects the increase in the industry's output due to technical change—i.e., the difference between what the industry's output would be with and without technical change for a given amount of investment.

- This TFP measure is not, however, a measure of technical change *per se* but only of what Jorgenson and Griliches call free gifts—returns over and above those needed to recover the costs of the R&D that created the innovation. This is easiest to see in case (4) where the rise in output is caused by technical change but the change in TFP is zero. This is because the rise in output would have been the same if there had been an equivalent amount of capital accumulation but no technical change. *Thus, zero TFP does not mean zero technical change, only that investing in R&D has the same effect on income as investing in existing technologies (investment with no technical change).*

- In the cases considered here, the fact that the technical change is embodied in capital goods does not lead to an understatement of TFP, provided that TFP is understood to measure J&G's free gifts, and not technical change *per se*.
- Free gifts occur in two frequently observed cases of technological change: first, new machinery increases the efficiency of both labour and capital; second new machinery increases the efficiency of labour but not of capital. In these cases, there will be positive TFP numbers even though the value of the R&D is capitalized in the price of the final product. It is not correct, therefore, that TFP should be zero when the costs of developing new technologies are fully allowed for. This is only true when R&D yields the same additions to output as would have occurred if the equivalent resources had been devoted to making new capital with existing technology. This agrees with Jorgenson's position. Our important difference is only to point out that in some types of technological change, commonly found in practice (and commonly used in theoretical literature), the requirement is not fulfilled so that TFP measures will be positive.
- These examples suggest that the changes in TFP depend on the sort of technological change that is being experienced. To reiterate: technological change that acts as if it raises both efficiency parameters leads to large changes in TFP; technological change that acts as if it raises the efficiency parameter attached to labour leads to changes in TFP equal to the increase in output weighted by labour's share in total cost; technological change that acts as if it raises the efficiency parameter attached to capital leads to no change in TFP. This kind of technological change is empirically indistinguishable from an accumulation of the same amount of capital with no technological change.

Natural resources made explicit

Background

Following Solow (1957), economists typically define the stock of physical capital used in their growth models to include the stock of natural resources, land, minerals, forests etc. Almost invariably, however, everything that is then done is appropriate for physical and human capital but takes no account of the specific problems of natural resources. For example, although the stocks of plant and equipment can be increased more or less without limit, the stocks of arable land and mineral resources are constrained within fairly tight limits.

The shortcomings of this treatment of resources can be seen in the contrast between two positions. The first is the prediction derived from the standard formulation in equation (III.2) below that measured capital and labour could have been increased at a common steady rate from 1900 to 2000 with *constant technology* and no change in living standards. The second is the argument given in section I of this paper (see "Holding Technology Constant") that such an event would have had catastrophic effects on living standards. To reconcile these conflicting positions, we need to recognise that the capital that would need to grow would include such resource inputs as acres of agricultural land, quantities of mineral resources, available "waste disposal" ecosystems, supplies of fresh water, and a host of other things that the standard measurements of capital ignore. (*Since technology is assumed to be constant, this growth cannot be in efficiency of use due to new techniques.*)

Since the evolution of stocks of capital often obey laws that differ from those that govern the evolution of the stocks of natural resources, it is conceptually useful to separate the stock of natural resources from the stock of created physical capital. Natural resources are used up in the process of production and some, such as petroleum, cannot be replaced while others, such as trees and the fertility of land, typically (although not invariably) can be. Most renewable resources, such as the air and water that remove pollution and nurture fishing stocks, are naturally renewable up to some maximum rate of exploitation. Above that rate, human help is needed to sustain the yield. Above some yet higher rate of exploitation, the stocks may deteriorate even in spite of human intervention. To account for these stocks, we need to deduct the natural resources used in current production, and treat as gross capital investment the amounts spent in maintaining existing resources (e.g., the productivity of land) restoring those used up (e.g., reforestation) and discovering new supplies (e.g., mineral exploration).

Importantly, technological advance alters the economic value of existing natural resources. Some values may be lowered, such as when the invention of the electric motor and the internal combustion engine lowered the value of coal reserves. Other values may be greatly enhanced (sometimes starting from a base of zero value), such as when the introduction of the gasoline engine and the automobile greatly increased the value of petroleum reserves. These changes in resource values are sometimes consciously created by technological change, as when methods of using low-grade iron ore and tailings from previous operations, were invented. At other times, they are the unconscious result of technological advances pursued for other purposes, as when the internal combustion engine, which started out as a stationary engine driven by coal gas, finally settled on petroleum products as its most efficient fuel.

The absence of explicit resource inputs from the neo-classical growth model, poses no problem as far as income goes because all of the value of consumed resources must show up as income for the labour and capital services involved in extracting and processing them.

Modelling resources

To illustrate some of the problems associated with the omission of natural resources, let the underlying production function be:

(III.2) $Y = AK^{\alpha}L^{\beta}R^{\delta} \qquad \alpha+\beta+\delta = 1$

where *K* is produced capital, *L* is labour and *R* is natural resources, agricultural land, forests, minerals, air, water, etc. [17]

Now let *K* and *L* increase at a constant rate μ. If nothing else happens, there will be diminishing returns as more labour and capital are applied to a given resource base. Output will be increasing at the rate $(\alpha+\beta)\mu< \mu$ and so per capita real income will be shrinking.

Now reallocate some of the capital formation to creating technological change in the resource industries. Assume that the new technologies are resource saving (as the evidence cited earlier in this paper shows that they are on average—although not, of course, in every individual case). Further assume that, as a result of the R&D, resources are also growing at the rate μ, measured in *efficiency units*.[18] Total income, capital and labour are now all growing at the rate μ while per capita income is no longer falling. However, if we measure *R* in physical rather than efficiency units, *R* will be constant, while *A* will be growing at the rate $\delta\mu$.

Now assume that we let (III.2) generate data over some long period of time. We then fit a constant returns to scale production function to it with *R* excluded and *K* defined conventionally to exclude stocks of natural resources. We will then get a perfect fit with the equation:

(III.3) $Y = BK^{\varepsilon}L^{1-\varepsilon}$

where $B = AR^{\delta}$ and ε will exceed α by the share of capital in the costs of producing resources and $1-\varepsilon$ will exceed β by the share of wages in resources costs. Thus, we can take the growth process that is actually being partially driven by "resource enhancing technological change" and fully "explain" it in terms of a constant returns production function containing only two inputs, *K* and *L*, and with an unchanged productivity parameter, *A*. So measured TFP will be zero and all the increase in output will be ascribed to increases in measured labour and measured capital.

We conclude that because resources are not specifically modelled in the neoclassical growth model, much of the substantial amount of technical change that goes into increasing the productivity of given natural resources will show up as measured increases in capital and labour rather than as shifts in the production function.

Notice that this result conflicts with Griliches' conclusion about the effect of unrecorded inputs. In his discussion of errors in measuring TFP (see above), he argued, not implausibly, that an increase in any unmeasured input would raise output without raising measured costs and so would add to measured TFP. *The present discussion shows that technological change that increases the efficiency of an unmeasured input may show up*

[17] We could add human capital but this would not affect the argument that depends only on having one major unmeasured input.

[18] The long-term evidence suggests that this is a better characterization of the resource sector than that given by the Hotelling rule where resource prices rise at a rate equal to the real interest rate.

as an increase in the quantity of other measured inputs and hence leave TFP understated. This is yet another reason why changes in TFP do not measure technological change.[19]

Timing of cost reductions

Assume now that there is genuine technological advance that allows cost reductions on all future output and that the present value of the cost reduction greatly exceeds the R&D costs of developing the technology. Thus, we have a productivity increase by J&G's stringent definitions, as well as by more inclusive definitions. How much of this will show up in TFP measures? It depends on timing.

If an established industry develops a method of producing at lower cost, which it embodies in new capital as its old capital wears out, there will be a sustained increase in measured productivity during the period when the old capital is being replaced by the new, which may last decades. In contrast, if a new industry cuts its costs dramatically when it is small and then expands greatly, there will be little change in measured productivity. Yet at the end of the process, a mature industry is dominated by some new technology in both cases. *The technological dynamism is similar but the TFP experience is radically different*.

Something like the former case occurred when electricity replaced steam in established manufacturing industries in a process that took several decades. Something like the latter case occurred in the new automobile industry when Henry Ford drastically cut production costs, after which his output rose dramatically. (Costs continued to fall as output rose to its peak but many of the most dramatic reductions came at the early stages when output was relatively small.)

An example

To illustrate the effects of timing, we stylize by assuming that all the cost reductions happened before most of the sales increases occurred.

TABLE 1

Period	Resource cost	Output	Unit cost
1	100	10	10
2	200	100	2
3	40,000	20,000	2

Table 2 shows the resulting changes in TFP.[20]

[19] This is a caution for macro modellers and those who calculate TFP from aggregate data. However, industry studies that include resources as inputs of downstream industries and as outputs of basic industries may catch these productivity gains by showing a *ceteris paribus* decline in resource inputs. (Although it is unclear whether or not this gain might be washed out when aggregating over all the sectors.)

TABLE 2[21]

Change in period	%change in inputs	%change in output	% change in TFP
1-2	66.7%	163.6%	96.9%
2-3	198%	198%	Zero

Since all of the productivity gains come when output is low, the change will get little weight in an economy-wide index and so there will be little observable change in total factor productivity at the economy-wide level. This timing problem will not affect an industry study, which will correctly show large efficiency gains in the early period and lower ones later on. But those who study macro productivity figures will never see the significant changes that are taking place.

New products and processes more generally

This is not an uncommon situation. New products and processes that are being developed rapidly are often produced by a large number of small competing firms. This is often the time when costs are reduced drastically. Once the technology of the product becomes more settled, the time series of unit costs begins to level out while expansion of sales is rapid. (Early on, the product is not reliable enough for mass use and customers take time to learn to use it.) To the extent that cost reductions come while sales are small, macro-level TFP will hardly be affected during the period of cost reduction, while both micro and macro TFP gains will be small during the period of rapid expansion of sales. To the extent that cost reductions come once sales are large, as with established industries that are introducing cost saving innovations, TFP gains will be large—assuming that the cost reductions are of the sort that will show up as TFP increases.[22]

Our analysis is also a stylization of the findings of Crafts and Harley for the early stages of the Industrial Revolution. The early textile sector was sufficiently small that gains in productivity in the first factories were not reflected in significant changes in overall national productivity. Nonetheless, fundamental changes were occurring in the organisation of British industry. These changes revolutionized all of manufacturing as they spread throughout the economy over the next half-century. The low national TFP figures (and low national figures for national labour productivity, Y/L) are correctly used

[20] The table also shows that one must be careful in interpreting the magnitude of TFP changes. In this case, unit costs fall to 1/5th of their original value but TFP only (approximately) doubles.

[21] All percentage changes are calculated on a base of the mean of the two period's values.

[22] A chain-linked index may pick up some of this. How much it picks up depends on the length of the chaining and of the gap between the fall in costs and the rise in sales. We know from technological history, however, that a fall in the price of one input that is widely used in many industries may cause redesign of many other capital goods. Such redesign and related innovations may stretch over decades, as will the increases in the output of the intermediate inputs whose price reduction initiated the whole series of downstream adjustments.

when they are interpreted as showing that current changes were having little effect on the whole economy. They are incorrectly used when they are interpreted as denying that a technological revolution was taking place.

Conclusion

This example is not intended to show any conceptual flaw in TFP. It is intended to show that major technological changes that enable the transformation of an industry based on a new technology that was originally introduced when the industry was young and small (as so often happens in reality) will produce only small TFP gains. In contrast, similar technological changes introduced into an already large established industry will show large TFP gains.

Some users of TFP measures have told us that what we have illustrated is just what a TFP measure is intended to do. There were few productivity increases as the industry in our illustration grew to maturity so TFP numbers should be small. Of course, one can measure anything one wants and those who fully understand the measurements will not draw erroneous conclusions from them. But the quotations at the beginning of this paper suggest that people do draw erroneous conclusions fro m low TFP numbers that arise because of the sequence just illustrated. *Whatever else it shows, the example illustrates how misleading it is to take TFP measures as indications of the importance played by technological change in economic growth and social and economic transformation.* Figures in the Tables, which are stylizations of what actually happened, yield small aggregate TFP gains emanating from the automobile industry in the period 1908-1928 (the lifetime of the model T). And yet without the technological innovations introduced by Henry Ford, the North American economy and society (and later the European) would not have been dominated by the automobile in the way it was, and still is. Technology really mattered, although on our stylized assumptions, measured macro TFP gains would have been less than dramatic.

A better measure

Of course, there is no “true” measure, but *a better measure of the impact of technological change, even at the industry level, would be the difference between the resources it actually took to manufacture the mature industry’s output and the resources that it would have taken if the earlier technologies had been used.* In the example in the above table, the resource cost is a mere one fifth of what it would have been if the older technology had been used—a very large saving indeed! In the case of cars, the comparison would have been the cost of producing the 1928 production of the model T at 1908 costs. Of course this volume of cars would never have been built if costs had remained at their 1908 level, *but the calculation shows the power of technological change in making possible something of great benefit, even though no major changes in economy-wide total factor productivity need have been recorded in the process.*

Aggregation out of static equilibrium

When we have appropriate *quantity* measures so that a measure of each type of input is available at whatever level of aggregation is being used, percentage changes can be calculated, weighted, and summed to get overall percentage changes in inputs. In the standard indexes of TFP, the coefficients on each type of input that determine relative shares are the appropriate expenditure weights—given all the usual assumptions.

Two assumptions are critical. The first is that marginal productivities are not changing, a sufficient condition for which is that the economy be on an equilibrium growth path. The second is that the marginal products of any one factor of production are equated in all of its uses. In this section we consider the latter assumption.

The equality of a factor's marginal products in all of its uses requires that a full competitive equilibrium be obtained, something that is manifestly untrue in many real world situations.

To see the importance of this, consider two concepts of equilibrium in a discrete time model. In the first, which we may call full equilibrium, all adjustments have been made and no agent wishes to alter his or her behaviour from period to period. In the second, which we may call transitional equilibrium, each agent does not wish to alter behaviour in the period in question but behaviour does alter from period to period.[23] To see the relevance of the distinction for our purposes, consider an economy with two sectors "agriculture", *X*, and "manufacturing", *Y*. The production technologies in the two are given by the following:

$$X = AK^{\alpha}L^{1-\alpha}$$

$$and$$

$$Y = BK^{\beta}L^{1-\beta},$$

$$A < B, \alpha < \beta$$

Initially the whole economy is in full equilibrium. Now let productivity rise in manufacturing by increasing the value of *B*. This increases the marginal productivity of *K* and *L* in *Y* production. Labour and capital move from *X* to *Y* until these values are once again equated. But in the real world, these adjustments may take much time, even generations, particularly if the increase in *B* is itself spread over time rather than occurring all at once. While this adjustment is occurring, statisticians will be measuring TFP during a time in which full equilibrium is not obtained.

There are various possible ways in which we may model the delays in adjustment, but one way is to assume that there are costs of movement for existing factors—e.g., selling a home and moving from the farm to the city—that exceed the difference between the wage (which equals the value of the marginal product) in *Y* and in *X*. Now no existing unit of labour will wish to move. New entrants are assumed to have zero movement costs and so will choose to enter either sector on the basis of current rates of pay. So as existing factors die or retire, new entrants will all chose to enter *Y* where earnings are higher than in *X*. Throughout this transitional period, transitional equilibrium exists because every agent is optimally adjusted to the circumstances he or she faces, but the allocation of resources changes period by period until the marginal product of *L* is the same in both *X*

[23] This is a distinction that Hicks long ago called full equilibrium and weekly equilibrium and which Archibald and Lipsey used to criticize Patinkin's analysis of monetary equilibrium.

and *Y. During the adjustment period, which may take decades ¾and may continue indefinitely if B is changing continually ¾the conditions for full equilibrium do not apply and unbiased aggregation to a macro production function is not possible. Importantly, relative shares do not then provide appropriate weights for aggregating percentage changes in labour forces.* Indeed, expenditure weights are not then free from contamination from marginal products and hence from assigning changes in productivity to changes in factor inputs.

The process we have just described is what was found in most industrial countries in the era of movement from farm to city that stretched from the mid 19th to the mid 20th centuries (with the details varying from country to country). Labour and capital was leaving the farm and moving to the cities where wages were higher. Full equilibrium was not obtained until the movement stopped sometime in the mid 20th century. During the transition, the conditions for unbiased aggregation were not fulfilled.

The following tables present a simple numerical example to illustrate the problems that arise. They show a reallocation of some factor from low to high return employment that, on a pure physical measure, would show no change in total physical inputs. However, in this example, the expenditure weighted Divisia index shows a 16.2% increase in inputs.[24] It is clear that if we change the value of the high productivity employment, we will change the weights and hence alter the computed change in aggregate input. For example, if the high productivity employment job had only paid $2 per unit of input, instead of $5, the weights would have been .182 and .818, making the overall weighted percentage change only 1.51%.

TABLE 3: FACTOR X'S MARGINAL PRODUCT IS HIGH IN SECTOR B RELATIVE TO WHAT IT IS IN SECTOR A

BASE PERIOD					**GIVEN PERIOD**		
Sector	Quantity of input X	Unit price (MP)	Value	Weight	Quantity of input X	% change	Weighted change
A	90	$1	$90	.643	80	-11.76	-7.561
B	10	$5	$50	.357	20	66.67	23.8
Total	100		$140	1.000	100		**16.239%**

{All percentage changes use the average of the two period's values as the base.)

[24] If we use the average of the base and given period expenditure weights, we increase the calculated increase in inputs to 24%, while the use of given year weights increases the value yet further to 32%.

TABLE 4: SECTOR B HAS A LOWER MARGINAL PRODUCT RELATIVE TO A THAN IN TABLE 3.

BASE PERIOD					**GIVEN PERIOD**		
Sector	Quantity of input X	Unit price (MP)	Value	Weight	Quantity of input X	% change	Weighted change
A	90	$1	$90	.818	80	-11.76	-9.62
B	10	$2	$20	.182	20	66.67	11.13
Total	100		$110	1.000	100		**1.51%**

The problem arises here because wages and marginal products are not in full equilibrium. Anything that introduces a lag in the adjustment of factors to earnings differentials will give rise to such a situation. There is nothing in this that violates the neoclassical assumption of full rationality and continual equilibrium; all that is required is that there be rational reasons why all adjustments do not take place instantaneously, the costs of movement and retraining being obvious examples. In these circumstances, weighting percentage changes in labour (or any other factor) by its relative shares in the various occupations (whether the weights are based on base year or given year expenditures, or and average of the two) will assign changes in productivity and in the degree of industrialisation to changes in the quantity of labour (and the quantities of other factors similarly affected). This is the type of lagged adjustment we see when a new technology starts to spread widely across the economy and may well be one source of slowdown in measured productivity figures.

There may be circumstances in which it is desirable to correct for the quality of labour, as is done in the above tables. But this is not appropriate when one is trying to separate changes in physical inputs from changes in productivity. Assume, for example, that Table 3 refers to one country, *S*, and Table 4 to another country, *T*. Both have the same reallocation of physical labour, but *S* gains more from it because it has technology that is superior to *T*'s. But if we do a Divisia index of the changes in TFP, we will assign much of the difference to a larger movement of labour in *S* compared with *T*.

Identical considerations apply to capital whenever costs of, or other impediments to, movement allow the marginal product of capital to remain lower in a contracting sector than in an expanding sector for some time. This will occur, for example, in a developing country whenever capital is moving over time from lower productivity, technologically stagnant sectors to higher productivity, technological dynamic sectors, TFP calculations with constant expenditure weights will assign part of the increase in productivity to an increase in the quantity of capital.

This illustrates that the calculated index of cost changes weighted by expenditure shares reflects the marginal productivities of the inputs, which tends to distort TFP calculations *when full equilibrium does not obtain (which, given lagged adjustments, is more or less all the time).*

Distinguishing and measuring technological knowledge and human capital

We need to distinguish three concepts that are not always kept distinct, technological knowledge, human capital and physical capital. These are shown schematically in Figure 1. (All Figures are at the end of the paper, before the Appendix.) Technological knowledge may be thought of as the accumulation of knowledge created by R&D (broadly understood). This knowledge then becomes embodied in physical capital when capital goods are manufactured, and in human capital when people are trained formally or acquire (explicit and tacit) knowledge on the job.

Appropriate depreciation

The concept of depreciation is similar for both types of capital: machines wear out while people retire or die. New physical capital needs to be created and young people trained. However, depreciation is significantly different for technical knowledge, it does not wear out and only occasionally is it lost.

Obsolescence is similar for all three of these types of capital. Machines become obsolete either when new machines can produce the same services at lower opportunity cost or when new product developments remove the demand for their services. Human capital becomes obsolete when the knowledge it embodies is no longer valuable. Obsolescence of technological knowledge does not follow any one simple pattern. At one extreme, knowledge of how to make some specific new product may become obsolete in a matter of a few years (even a few months) as the product is replaced by some preferred alternative. At the other extreme, some technological knowledge has a useful lifetime measured in centuries, even millennia—e.g., the wheel, the lever, the screw, and the dynamo. So for many purposes, it is important to do what is often not done in theoretical models and in measurement: distinguish technological knowledge from human capital and acknowledge that different depreciation-obsolescence rates should be applied to each. (Of course, it is still a vast oversimplification to define an aggregate stock of each and then apply a single rate of depreciation to each stock.)

Accumulation

Now consider the accumulation of human capital. As observed earlier, to estimate the effects of accumulating more human capital while holding technology constant, we need to think of educating more people to the full level of knowledge existing at some base period, e.g., 1900. Any knowledge that we give them about new product, process, and organisational technologies that were generated after the base period is the embodiment of technological advance.

In practice, the accumulation of human capital over time necessarily entails the embodiment of new technological knowledge in the same way as does physical capital. It is conceptually difficult, therefore, to separate what should be regarded as the output effects of new human capital from those of new technological knowledge. We are more productive today than we were 100 years ago mainly because we know a lot more about how to produce, what to produce, and how to organize our activities. If we include these types of knowledge as human capital, we measure the effects of technological change as increases in human capital. Whether we do this or not is largely a matter of taste, not of right or wrong. *If we do so, however, we are not justified in concluding that technological change accounts for little of the observed increases in output just because increases in inputs, including increased human capital, can do so statistically*.

One common measure of human capital is the amount of time spent in education. This measures the opportunity cost in terms of hours of labour input. Of course there is a problem in estimating the amount of time spent in informal education: on the job, and other forms of non-formal education, which is a significant number and one that can change as a proportion of total time spent in all forms of education. This is no easy matter. Even if this could be done accurately, there is still the problem of what kind of knowledge is embodied in human capital.

When we use time series data, we find that people typically spend more time in school today than in say 1900. This is partly because there is more to learn for all levels of entry into the labour force from "unskilled" labour, to trades persons, to professionals. Today's contribution of human capital to output would be much smaller than it actually is if the longer time in school was spent in learning only the knowledge available in 1900 compared with learning today's knowledge. Yet the extent of this difference does not measure the results of the accumulation of "pure" human capital; instead it measures the results of embodying new technological knowledge in human capital. [25]

Similar problems of disentangling the influence of pure human capital from the influence of the technological knowledge that is embodied in it arise when dealing with cross sectional data. To make useful cross section comparisons, we need to understand not only *how much* is known but also *what* is known. For example, eight years in school studying Marxist philosophy and the sayings of chairman Mao would produce less valuable human capital than eight years studying the "three Rs". In another paper, we have studied an important example of this point drawn from the history of economic growth. We observe that Kenneth Pomeranz (2000) has shown that in the 18th century, the Chinese had a high level of literacy, probably equal to the level of literacy in European countries. But a quantity of human capital similar to that of Europeans does not imply that the Chinese were on the verge of the Industrial Revolution that the Europeans created in the late 18th and early 19th century. We put it this way:

> "The Chinese levels of literacy only provided the capacity to accumulate human capital. The human capital that was accumulated in China was mainly the knowledge laid down in the profoundly non-scientific and non-analytical official civil service entry examinations to

[25] An alternative measure is the opportunity cost of education measured in output fo rgone. But the opportunity cost of an hour in education is just the marginal product of labour, which as we have already observed, contains the TFP constant *A* in it in a macro model (and is influenced by the productivity of labour in a micro model). So the opportunity cost of labour is higher in the US than in Bangladesh partly because the technological knowledge in use is higher in the US than in Bangladesh.

> which, as we have noted, most education was geared. Although Europeans may have had no more receptor capacity in term of literacy rates than the Chinese, they were acquiring a very different body of human capital. Science was all the rage in Britain. Educated persons were striving, and by and large succeeding, in acquiring an understanding of Newtonian mechanics—the basis of understanding most of the science and technology on which the First Industrial Revolution was based. At the same time, the Chinese were learning about other, non-mechanical, things. (Neoclassical growth models treat human capital as a homogeneous input, but, as this example illustrates, the nature of what is known is at least as important as the quantity.)" (Lipsey & Bekar 2000)

Correctly measuring the quantity of human capital and allowing for variations in its quality are important, particularly for TFP studies based on a single macro production function, which usually include a single index number of human capital as an input. In practice, any measure that is used cannot separate the accumulation of "pure" human capital from the accumulation of the technological knowledge that it embodies.[26]

R&D, the capital stock, and output

Since practices differ somewhat, we will look at what actually happens to R&D in the Canadian national accounts. On the input side, R&D is recorded as a current cost. It has no output counterpart until its fruits are used to alter the output of final goods.

Probably the best measure of the increase in production associated with R&D would be the discounted present value of the additional output that it is expected to produce when its results become embodied in new product and process technologies. But this expected income stream is not what is typically measured in practice where R&D is strictly regarded as a cost with no direct output.

It follows, for example, that if an established Canadian-based firm shifts resources from making machines into R&D to design better machines, it will record a fall in output with no change in input costs and hence, ceteris paribus, a reduction in its TFP. Presumably, however, it switched resources because they could create *more* value in R&D than in production. Whatever else we may think about the desirability of having such a characteristic in TFP measures, the resulting change in TFP emphatically does not measure any change in technology. Instead, it measures a change in the allocation of

[26] Here is another illustration of the kind of error that can arise from the inability to distinguish pure human capital from the technological knowledge that it embodies. Consider the following two countries. Country A has an elaborate set of technology enhancing policies while *B* has none. Years of schooling are higher in *A* than in *B* because there is more technological knowledge to be learned in *A* than in *B*. If we ascribe the superiority of *A*'s productivity over *B*'s to a higher quantity of human capital, we are measuring differences in available technologies as differences in human capital. As already observed, there is no one "true" way to define our variables, but when human capital is defined as above, measures that produce similar TFP residuals and account for output differences by difference in the input of human capital cannot be used to argue that *A*'s superior technology enhancing policies are ineffective. (Arguments similar to this can be found in the literature.)

resources in response to a judgement that the payoff to R&D has increased. This is what the economy has been doing over recent times as an increasing proportion of expenditures go into product design and other R&D compared with direct production costs.

A start-up firm that does only R&D in one year will have its output valued at cost and record an equal negative profit, since it has no sales. Thus, by definition, not only will it show a negative contribution to TFP, it will show no contribution to current output. In fact, of course, it may be contributing to technological dynamism by producing valuable new product and process technologies that are embodied in new patents.

If the patent produced by the R&D is sold abroad, this is recorded as a capital transfer. No income is ever recorded and hence there is no TFP gain at any point in the process. This is also the case if the start up firm is itself sold to a foreign multinational. Many Canadian firms do just this, engaging in start up behaviour and then selling out to foreign multinationals, making the return on their R&D expenditures in terms of the sale price of the company. (Indeed, Canadian tax advantages given to small firms encourage such activities.) None of this value-creating activity, often in the "new economy," will show up as income or as changes in TFP. If the patent is sold to another domestic firm, this is regarded as a capital transfer and there is no possible effect on TFP until after the new technology is put to use.

Currently, much R&D is going into developing new technologies and new products that will pay off in the future. Any increase in R&D expenditure (whether or not it is caught in national income) has no measured counterpart in increased output. This continues until the R&D produces results that are not just embodied in valuable property rights but in actually processes that begin to raise output faster than current costs.

So in these respects, TFP measures nothing systematic concerning the value created by R&D until the new technologies are used to reduce costs or increase the production of final goods and services. Furthermore, there is a potential for getting misleading TFP measures as the economy switches more and more from investment in producing hardware to investment in producing ideas.

Human capital

Similar considerations apply when there is much learning by doing and learning by using associated with a major new technology. The associated costs will show up as increases in current costs with no corresponding increase in present output. Such expenditures might be better regarded as capital investments in making the new technologies work. Of course, it would be impossible to isolate them and treat them as such, but it is worth noting that they give rise to measured increases in costs without measured increases in current output.

The slowness with which a new GPT makes its effects felt on productivity, contrasted with the speed with which it raises costs, is a general case in point. Paul David's famous article (David 1991) on the slowness with which electricity made its effects felt provides an excellent illustration of this general point. *These differential rates provide one important reason why, correctly measured as current flows, productivity does not accelerate, and may even slow, when a radically new technology is spreading through* *the economy. Much up front investment in learning by doing and using and the adaptation of an evolving GPT to new uses is taking place with delayed effects on output.*

IV. EXTERNALITIES, TECHNOLOGICAL COMPLEMENTARITIES AND EQUILIBRIUM

In this part, we introduce the concept of technological complementarities, which derive from the complex set of interdependencies that characterize any technology system. Our analysis leads to three important conclusions. (1) Externalities as conventionally defined do not capture the external effects that new technologies have on firms industries and sectors beyond those where they originated—instead these are captured by a much wider concept called technological complementarities. (2) Technological complementarities are the vehicle for sustaining the growth process—a growth that will at various historical periods be faster or slower depending on the nature of the complementarities associated with successive major technologies. (3) The effects of technological change must be measured by a counterfactual experiment and are not, therefore, correctly measured by total factor productivity.

Growth through a succession of GPTs

We begin by saying a little more about technologies and technological change. The overall technology system of any growing economy evolves along a path that includes both small incremental improvements and occasional jumps. To distinguish these, investigators often define two categories. An innovation is *incremental* if it is an improvement to an existing technology. An innovation is *radical* if it could not have evolved through incremental improvements in the technology that it displaces—e.g., artificial fabrics could not have evolved by incremental improvements out of the natural fabrics that they displaced in many uses.

An extreme form of radical innovation is called a general purpose technology (GPT). 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 complementarities* 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.[27] The steam engine, the dynamo and the internal combustion engine are examples of major GPTs in the field of energy generation.

An important characteristic of a GPT is that it creates a new research program for inventions and innovations that use the GPT either directly or indirectly—indirectly in the sense of using other technologies that in turn use the GPT. As the research program

[27] For a detailed consideration of these characteristics and a development of the definition that follows in the text see Lipsey, Bekar and Carlaw 1998 a.

proceeds, new opportunities open up exponentially. Then as the GPT matures, the number of new opportunities created per unit of time may begin to fall, causing a reduction in the returns to further investment in invention and innovation based on that GPT. A stylized version of the results of this evolution is a logistic curve, which has realised investment opportunities associated with a given GPT on the vertical axis and time on the horizontal axis. Note that the curve does not flatten at its upper end as a result of diminishing returns to the growing capital stock with technology constant but rather because of a reduction in the rate of creation of new investment opportunities as a GPT matures, leading to a fall in the rate of realisation of new derivative innovations.

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 programs for applying electricity were very different from those for applying the steam engine. (The view expresses in this paragraph is formalised in Section V.)

In contrast, most aggregate growth models apply a single macro production function over time periods that include several successive GPTs. Holding the labour supply constant in such a function, there is a continuously declining marginal product of physical and human capital that is only increased by continuing gradual increases in the productivity constant, *A*. For balanced growth, when the non-labour inputs are increasing at some constant rate, the efficiency parameter must be increasing at a rate sufficient to produce Harrod neutral balanced growth.[28]

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 GTPs, 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 characterize individual firms, industries and sometimes whole sectors.

[28] For example if capital is growing at the rate r in equation II.1 with labour held constant, A must be growing at the rate $(1-\beta)r = \alpha r$.

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 generalizations 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.

Definitions

Next, we need to examine carefully the concepts associated with the external effects of technological change.

Externalities

As it is normally understood, we can define an externality as an unpaid-for effect conferred by the actions of one set of agents on another set of agents not involved in the first set's activities. We refer to these as the "initiating agents" and the "affected agents" respectively, and we confine ourselves to situations in which the affected agents are impacted favourably. A favourable production effect is usually analysed as an unpaid-for input into the affected agents' production function that is created by the initiating agents' activities. The formalisation of one such case is as follows[29]:

(IV. 1) $A = ax_a^{\alpha} y_a^{\beta}$

(IV. 2) $B = bx_b^{g} y_b^{d} A^{f}$

Where A and B are outputs of two products, x_a, x_b, y_a and y_b are two normal inputs used by A and B, and *A's* output occurs with a positive partial derivative in the production function for *B*. In a standard GE model with tastes and technologies constant, externalities arise in the following ways: (i) between producer and producer as in the above equations, (ii) between consumer and consumer as when some aspect of one consumer's consumption appears in another's utility function, (iii) between consumers and producers as when some activity of one appears in the utility or production function of the other.

If the output of *A* rises in the above case, then the output of *B* will also rise with no increase in *B*'s paid-for inputs. When output of B is explained using a production function with only x and y as inputs, the effect of the externality will show up in the productivity constant *b*. Thus changes in *B*'s TFP will be positively correlated with changes in *A*'s output. The existence of this externality could be tested for by looking for a positive correlation between *B*'s TFP and *A*'s output.

[29] In other cases, the quantities of one or more of the inputs used by A may appear in B's production function.

There are three issues that need to concern us with respect to externalities as they are usually defined. They are timeless; they concern continuous actions; and they do not cover an important type of interaction.

First, all standard definitions say nothing about when the externality creating action took place. Since the standard analysis is usually concerned with how externalities upset optimality in a timeless model, the action is clearly assumed to be a contemporaneous, and a continuing one that is capable of being altered. So an action that has been completed and is irreversible presumably does not create a current externality in any operative sense. Although agents making current decisions are affected by the past actions of many other agents, these bygones do not upset current optimality conditions.

Second, inventions and innovations are one-time discrete events that do not fall into the framework of constant-technology GE models. When economists seek to study the externalities associated with technological change, they typically get around these problems by considering R&D. This gets around the first problem because R&D is a continuous variable that is assumed to create a continuous stream of value for the affected agents that they do not pay for (an externality). The second problem is usually circumvented by modelling R&D as in input into all agent's production functions rather than as shifting their production functions as it typically does.[30]

Third, an important class of interactions occurs when an invention in industry *A* creates an opportunity for agents in industry *B* to make another invention that incorporates, or in some way relies on, the invention in *A*.[31] The action in *A* does not directly shift the production function for *B*. Instead it creates an opportunity for agents in industry *B* to conduct R&D to invent a new technology and invest in installing it. This will eventually alter production functions in industry *B*, either by altering the relation between inputs and an unchanged product or by altering the nature of the product itself. This important type of relation is not necessarily an externality in the sense of giving something freely to agents in *B* that they would be willing to pay for. The R&D costs of developing the new innovation in industry *B* may just be covered by the income earned from the new production function—a possibility that is stressed in the J&G view.

This discussion lead us to distinguish between two types of externality.

An externality as an unpaid-for effect conferred by the actions of one set of agents (the initiating agents) on another set of agents (the affected agents) not involved in the first set's activity.

- A standard externality occurs when the actions are those that can be made in a standard Arrow-Debreu type GE model with constant tastes and technology.

[30] This is yet another problem of aggregation that is ignored in standard treatments. At the micro level, R&D shifts production functions and aggregation would have to be over the set of differently shifting functions. In practice, an aggregated production function is merely *assumed* with R&D as an input in addition to the other traditional ones. Thus no distinction is made between input activities that alter outputs within given micro production functions and those that shift these functions.

[31] We use the term invention and innovation interchangeably here because the distinction, although important in many contexts, is not important for our present argument.

- A technological externality occurs when the action of initiating agents is to make a technological change that freely provides an opportunity for the affected agents to make further technological changes, an opportunity for which they would have been willing to pay.[32]

Notice that the standard treatment of R&D discussed above has the effect of turning a technological externality analytically into a standard externality.

Definition 1.b covers a wide range of opportunities. Here are two key examples. First, two agents may be striving for the same advance. Agent 1's solution to the problem enters the public domain and saves agent 2 from further R&D expense. Agent 1 confers an externality on agent 2 for which the latter would have been willing to pay up to an amount equal to the expected cost of carrying his own research program to completion. If agent 2 had been expecting to earn a normal rate of return on his activity on the assumption that he would have to pay the full R&D cost himself, then the free gift of agent 1's results should produce additional profits which will show up in 2's TFP.

Second, agent 1 may make a technological advance that agent 2 can use as a basis for making further advances. *If* agent 2 makes more than a normal rate of return on its innovative activity, she would have been prepared to pay agent 1 for giving her that possibility. The extra amount should also show up in 2's TFP.

Because innovations take place in real time, this concept of a technological externality becomes far-reaching as soon as we leave timeless value theory. Agents continually enjoy the free use of technologies for which they would be prepared to pay something. This includes all previous technological advances that are now in the public domain and still in use. Any one 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 newly invented technologies 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 PCs if there were no electricity. Many agents would pay if the alternative were to have to do without one or more of these existing technologies. Thus the economic durability of many basic technologies gives rise to a large list of such technological externalities.

[32] These pure cases are all that we need for present purposes. However, cross cases are also conceivable. First, an alteration in a GE variable by the initiating agents could provide the opportunity for a technological change to be made by the receiving agents. 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 pure profits 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 may provide the opportunity for profitable substitution of inputs by the affected agents. This case is also common, occurring whenever new technologies alter the prices of existing inputs or outputs. This appears to the affected agents as a standard externality where a GE variable changes. Presumably, they would have been willing to pay something to the initiating agent to induce them to make the change in their technology.

Technological complementarities

Next we consider a wider class of interactions that we call technological complementarity, or technological interrelatedness or technological spillovers, all of which terms we use synonymously. We define these as *any situation in which decisions by one set of agents' (the initiating agents) with respect to their own technologies affect the value of another set of agents' (the affected agents) existing technologies and/or their opportunities for making further technological changes.*

Note that this definition does include the third class of interactions discussed above.

Figure 2 illustrates what is involved. The "upstream" initiating actions are shown as one of two main categories. First, the upstream innovation of some new technology, ΔT_0 in Figure 2, may change such market signals as prices and these may create incentives for technological change—either incremental changes in existing technologies or radical inventions of new technologies. For example, the development of the mobile internal combustion engine led to an increase in the demand for petroleum that provided the incentive for many innovations relating to its discovery, extraction and refining.

Second, a new technology, ΔT_1 in the figure, may make it possible to invent a host of other technologies without any change in a price or quantity signal. For example, the mobile internal combustion engine was critical to the invention of lighter-than-air craft where the engine's low weight to horsepower ratio was a decisive advantage over steam. In some cases, the new technologies may combine mainly with technologies already known, as when the upstream invention of the diesel motor and the dynamo allowed the replacement of steam engines with diesel electric engines in ships whose technology was already very sophisticated. In other cases, it creates possibilities for the invention of totally new technologies. Examples are the inventions of radio, electric lights, and TV, which were made possible by the invention of practical sources of electricity. Other examples of such technological complementarities arise when the invention of a new technology induces the redesign of existing technologies, sometimes because redesign is necessary for the new technology to work at all or, more often, for it to work efficiently.

The increase in the range of possibilities for new innovations created by a major new technology such as a GPT is not a once-for-all event. At first, the technology is crude with few uses so the enlargement of the range for new innovations is limited. But as the technology matures, it improves in efficiency and widens in application until it spreads through much of the economy. As this happen, the range for new innovations that it creates is continually enlarged. Each new innovation in the newly created range of possible innovations gives rise to new possibilities and so further enlarges the range. It is a characteristic of GPTs which start in crude form with limited applications, that the size of the space of new possibilities created by each successive round of innovations increases for a long time.

Clearly, these technological complementarities are much more pervasive than externalities. The Appendix gives five case studies that illustrate the many and varied ways in which one innovation creates the opportunity for others. Sometimes the initiating innovation is a piece of capital designed for specific purpose but found to have much wider applications, as with US machine tools in the 19th century (Appendix Case 1); sometimes the same thing happens with a production technique, as with the "American

System of Manufactures," originally developed for the production of guns and later applied to many other products (Appendix Case 2); sometimes one type of machine enables a whole set of subsequent new technologies, as with machine tools able to cut pre-hardened metals out of which came generalised mass production of most manufactured goods (Appendix Case 3); sometimes the initial innovation is itself a GPT, as was electricity (Appendix Case 4); sometimes the linkages are so complex that one GPT gives rise to a second GPT with its whole new set of complementarities, as when the computer and electricity enabled the Internet (Appendix Case 5).

It is important to note that the responses of one technology to a change in another cannot typically be modelled as the consequences 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. The consequent changes will typically take the form of new factors of production, new products, new sectors, and new production functions.[33] Consider two examples of this important point. First, the remodelling of the factory that came after the introduction of the electrical unit drive could not have resulted from a fall in the prices of steam power, even to zero. Second, the mass production factory that was introduced after the invention of machine tools that could cut hardened steel 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 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, externalities arising from new technological knowledge. In other words, if every final good and service was unique and had its own unique production process and unique set of intermediate goods with no overlap with those associated with other goods, new knowledge would be user specific without spillovers to, and externalities for, other goods and production processes. *Therefore, the existence of a technological complementarily is a necessary condition for the existence of a significant technological externality.*

Although technological complementarities are related to externalities, since their benefit is often not fully appropriated by the original innovators or others in the same industry, they are not identical with externalities. Two reasons are worth noting. First, the original innovators might be able to capture all of the extra value created by downstream complementarities. This would be most likely if the downstream applications of the new innovation were specific and localized. Second, there may be no extra value in the sense (already noted) that the benefits of the downstream innovations may just cover the R&D costs of developing them. Although there are no free lunches associated with either of these types of technological complementarity, they do provide downstream agents with opportunities that they did not previously have and that are worth taking up. *Both of these cases show that the existence of a technological complementarily is not a sufficient condition for the existence of a significant technological externality.*

[33] 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.

We have seen that technological complementarities are dependant on technologies being interdependent in such a way that one technological change impacts on many other existing and potential technologies both positively and negatively, giving rise to scope for improvements in some (incremental innovations) and as well as other wholly new innovations (radial innovations). What is critical for the distinction between externalities and complementarities is that these upstream innovations do not necessarily confer benefits (externalities) for which the downstream innovators would have paid since the downstream activities may pay for all the current resource inputs that they use and just earn the going rate of return on their investment.

In summary, let T be the class of all technological complementarities and E the class of all externalities. Then $T \cap E$ defines technological externalities; $T' \cap E$ defines standard externalities and $T \cap E'$ defines those technological complementarities that do not give rise to externalities as so are not covered by conventional measures of the external effects of technological change that seek to measure externalities.

Technological complementarities not externalities drive long term growth

To see 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. As we have observed earlier in this paper, there is only so much that can be done with Victorian technology. The scope for making production more capital intensive with no changes in technology is limited, since many technologies have input proportions built into them, proportions that can be varied only within quite narrow limits. Of course, there is more scope in some industries for increasing the capital/ labour ratio even with unchanged technology but there is little doubt that diminishing returns to capital would set in just as theory predicts.

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 widened 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

Economists have long attempted to measure the external effects of innovation. The attempts have usually implicitly assumed that these effects are all in the form of externalities that will show up in someone's TFP since they are unpaid-for benefits. A further assumption is that the external effects will show up in the outputs of the affected agents at the same time as they are influencing the output of the initiator. In other words, R&D can be entered into a production function in just the same way as any other input and its effects on output are subject to lags no longer than those attached to other inputs.

In this case we can write:

$A = aL_a^{\alpha}K_a^{\beta}R_a^{\gamma}R_b^{\delta}$

$B = bL_b^{\alpha}K_b^{\beta}R_b^{\gamma}R_a^{\delta}$

Where L_a, L_b, K_a and K_b are labour and capital used in *A* and *B*, R_a and R_b are R&D done by *A* and *B*, and the coefficient γ measures the effect on output of one's own R&D while the coefficient δ measures the effect of other people's R&D. These are heroic simplifications of the innovative process. First, it is assumed that R&D can be treated just like any other input in a flow production function whereas, in fact, its effects are long delayed, discontinuous and typically alter the production function by producing new products and new processes. This simplification is one of those intuitive leaps that pervade a literature that calls for highly formal modelling where that is possible, but accepts intuitive leaps where relations cannot be formally derived. Second, it is assumed that the external effects are felt contemporaneously with the original effects. In so far as innovations in *A* present a research program for *B* that may stretch over decades (as e.g., small unit drive electric motors presented the consumers durable industry with a program that lasted decades), we would not expect the effect to show up in any contemporaneous measure, nor would we expect it necessarily to provide the kind of current free lunches that show up in TFP.

Given that the appropriate assumptions are made, two methods have been used in attempts to locate the external effects of technological change. One method is to look for a correlation between the TFP 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. If there is a positive correlation, then this is taken as evidence that there is a positive spillover. This is a test for contemporaneous free gifts in the form of output not matched by paid inputs but not for all the other forms of technological complementarities that we have discussed above.[34]

A second method is to calculate the rate of return to R&D at various levels of aggregation. Here the R&D input to production is usually measured as the stock of R&D. The gross rate of return to R&D is then calculated as output divided by the input of R&D capital weighted by the estimated elasticity of output with respect to the R&D input (which requires treating R&D as a normal input into current output). In the standard Cobb-Douglas formulation this elasticity is just the estimated exponential parameter on

[34] Aside from all 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 firms output by $0.05. Let the two firms be engaged in active competition and let their R&D's be fairly well synchronized. 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.

the R&D input. (This is the coefficient on our R_a in equation IV.3 above with the coefficient on R_b set to zero.) An increase in the rate of return to R&D as the level of aggregation increases is taken as evidence of a positive R&D spillover. The social rate of return to R&D is taken as the difference between the rate measured in this way at an aggregate level and the sum of the rates measured at some disaggregated level.

Both of these methods have been used in many attempts and all have failed to indicate strong evidence of R&D spillovers at the aggregate level. This does not surprise us. Our own analysis strongly suggests that the measures are based on faulty conceptualisations of how the effects of R&D spread beyond those who actually conduct it.

We have shown that in our conceptualisation, what major GPTs, such as electricity, do—and occasionally a very specific innovation such as the machine tools that could cut hardened steel did—is to create opportunities for profitable investments in a large set of new product, process, and organizational 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,[35] 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, 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.*

Because they all interact with each other, there is no unique meaningful measure of the value of these complementarities. Consider an agent faced with the choice of doing without the first item on a list of all the 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. When technologies interact, cooperate with, and support each other, it is in principle impossible to measure the amount that each one separately contributes to the value created by the group as a whole.

We may get a clue as to how to proceed if we consider the simple case of the many new technologies that allow us to do what we were doing but do it more cheaply—the machine tools that cut metal faster and better are one example. More generally, we do things that were done 100 years ago at a small fraction of the labour and resource cost that was required with previous technologies and, hence, we can do more of them. In

[35] Of course, we have no doubt that there sometimes are enormous returns, sometimes 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.

these cases, the measure of the gain is the difference between the current cost of doing the job and what it would have cost to do the same job with the old technology—*which is not typically measured by changes in TFP* (as we saw for example in the case of the timing of innovation and of industry expansion considered in Section III). In other cases, 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 measuring the gains resulting from the replacement of the old by the new technologies. 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 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 organise production in ways that were technically impossible with steam. (This is what we meant when we said in the earlier section 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 been introduced.[36]

Consider the extreme case of no free lunches (case 3 at the beginning of section III). Technological advance could not be less productive than in case 3 (or it would be more profitable to invest in more of the old technology rather than in any of the new). In this lower-limit case, where returns to the new technology just cover the full costs of developing and producing it, the rate of return on the new technology will just equal the rate of return on the old at the outset. 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.

Consider an example. Let the current rate of return on investment in old technology in period t be 10%. Assume that if technology were held constant, diminishing returns to the accumulation of capital would drive that rate of return down by 0.1 point per year. Further assume that investment in bringing a new technology to market also earns a return of 10%—there are no free lunches. Investment in the new technology takes place and, as well as yielding 10% for its owners, it creates the opportunity for a new innovation in period t+1 that also yields 10%. In period t+1 investment in period t's new technology continues to yield 10%. 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 so it goes

[36] Typically, technologies have many uses and as a new technology evolves and improves it will challenge and eventually replace the old technology in uses where it has most advantage first, and in uses where it has the least advantage last. For example, steam took a century to drive sail out off all commercial uses, first on rivers, then on harbour tugs, then on cross Channel trips, then on transatlantic passenger travel, then on high value long distance freight and lastly on long distance bulk freight.

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. How 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 by 0.5% per year. By year *t* + 50, the gains will be the 5 percentage points between the constant returns of 10% to R&D and the 5% to which the returns to investing in period *t*'s technology would have shrunk by then.

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 counter factual 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 realized. 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 goods we now create. (Analogous arguments apply to changes in labour productivity.)

V. A NON-EQUILIBRIUM MODEL OF SUSTAINED GROWTH

In this section we introduce a simple growth model that captures some of the key characteristics of technological change discussed in Section IV. We use it to show that the growth process does not have to exhibit long run stationary equilibrium properties and that there is no need to assume increasing returns, or externalities or spillovers from R&D activity in order to sustain growth. 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 nonexistent.

We have developed a view of the growth process that is consistent with the micro observations of technological change. It has new GPTs arriving at widely dispersed times (usually there is more than one working through the system at any one time). The evolution of the GPT and its new derivative technologies creates a growing set of technological complementarities. Technological improvements, 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. Thus the impact of a single GPT on technological change may be roughly captured by a logistic curve. As GPTs follow each other, each one brings its own research program whose results follow its own logistic curve. Some programs are richer than others, so 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.

We now give the structural equations of a simplified model that captures at least some of these ideas. Working it out in full must be the subject of a separate paper. In the meantime, we note that the model assumes production functions in three sectors, each of which have decreasing returns to scale; and it explicitly models technological complementarities among different types of R&D. The economy comprises three sectors. One produces a homogeneous output of a final good; the second does applied R&D that applies the general purpose technology to specific uses, the third does fundamental R&D that seeks to generate new GPTs. All of these activities use a generic input (resources) to produce their respective outputs. A single allocation problem must be solved in every time period: how much of the resource to allocate to each sector?

Input: The resources devoted to the production of output, "applied" R&D, and "general purpose" R&D are respectively o_t, r_t and g_t. The resource constraint for the economy is:

$$R = o_t + r_t + g_t \, .$$

R does not vary through time.

Output: Output is defined as a single good that uses resources and the stock of capital created by applied R&D as inputs.

$$Y_t = A_t^{1-a} o_t^{a}$$

where $0 < a < 1$. A_t is the created capital stock resulting from applied R&D that makes resources more productive in producing output.

Applied R&D: This sector creates capital for the output sector in the following manner:

$$A_t = a_t + (1-e)A_{t-1}$$
$$a_t = B_t^{1-b} r_t^{b}$$

where $0 < b < 1$ and ε is the rate of obsolescence that applies to R&D. B_t is the stock of embodied technology coming from the general purpose R&D sector. Change to B_t make the resources devoted to the applied R&D sector more efficient in producing the stock of embodied technological change for the output sector.

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

We now introduce a further, general purpose, layer of R&D activity. This endogenizes the B_t term.

General purpose R&D: This sector creates a stock of embodied general purpose technology to be used in the applied R&D sector.

$$B_t = g_t + (1-d)B_{t-1}$$

$$g_t = \int_0^{g_t} f(x)dx$$

f(x) is a density function and, for a specific illustration, we have chosen f(x) to be the probability density function for the gamma distribution.[37]

$$f(x)=\frac{l^q}{\Gamma(q)}e^{-lx}x^{q-1}$$

with mean $q/_l$, and variance $q/_{l^2}$ and,

$$\Gamma(q)=\int_0^\infty z^{q-1}e^{-z}dz$$

is the gamma functionof θ for θ > 0, with

$$\Gamma(q)=(q-1)!$$

for all integers θ > 1. d is the rate of obsolescence, which applies to GPTs. We assume for simplicity that θ = 1. This allows us to work with the exponential distribution in our preliminary numerical analysis. Also, we assume that λ is a nature-given parameter of random size in a given time period. This conforms to the empirical evidence by making innovation an uncertain venture in terms of the size of outcome that might occur (with zero being a possibility).

The maximization problem is to allocate resources to maximize the expected present value of output in every time period:

$$\begin{aligned}&\max_{o,r,g}\sum_{t=0}^{\infty} r^t Y_t\\&s.t.\\&A_t = a_t + (1-e)A_{t-1}\\&B_t = g_t + (1-d)B_{t-1}\end{aligned}$$

(and also subject to the additional restrictions that we have placed on γ_t).

We believe that this model captures much of the essence of GPT-driven growth while being technically much simpler than the models found in Helpman (1999). Its salient characteristics are worthy of notice.

- If there is no investment in general purpose R&D so that *B* remains constant, the returns to applied R&D eventually reach zero and the model will settle into a stationary state with zero applied R&D.
- Because there is investment in fundamental R&D, the model is always in transition in the sense that the marginal products of resources allocated to the production processes are constantly changing. The particular pattern of growth will depend on initial conditions and the realization of the random parameter λ in every period.

[37] Any density function defined for a random variable that takes on only positive values would suffice here. The gamma is chosen because several different continuous distribution functions are direct transformations of it. Continuity is assumed for convenience only.

- The marginal productivities of resources will be changing as a result of two distinct forces. First, they will be changing in one direction as a function of diminishing marginal productivities dictated by the parameters of the production functions. Seconds, they will be changing in the other direction as a result of the stochastic process that dictates (in part) the size of new GPTs. When this second process produces a new GPT, it shifts the marginal productivity schedule of applied R&D. Then, when the output from applied R&D gets embodied in new capital for the output sector, the marginal product of resources use in output shifts out as well. R&D resources in the output sector are thus being made more productive from two sources: directly from the applied R&D sector that is creating capital for final output, and indirectly from the general purpose R&D sector that makes the resources devoted to the production of capital more efficient and thus the resource devoted to final output more efficient.
- Technological complementarities are explicit in the relationship between the applied R&D that creates capital and the fundamental R&D that makes the creation of capital more efficient. The resources that generate GPTs cause all subsequent allocations of resources to capital production to generate more output.

This model is still very simplified in that GPTs, applied R&D, and current production are each treated as a single aggregate sector, while a new GPT creates investment opportunities whose marginal product declines at the outset rather than first rising, then stabilizing then falling as the GPT matures. But what if does show is, first, how GPTs can sustain the growth process by presenting a new R&D program that shifts the R&D payoff function and, second, that neither increasing returns to scale, nor externalities, nor spillovers from R&D, are required to generate sustained growth. The need for these characteristics is an artefact a modeling process that makes stationary equilibrium a requirement of a growth process.[38] In a later paper we will develop this model in full.

VI. CONCLUSIONS

Here are some of the conclusions that the present study has reached and concerns that it has raised about TFP

- TFP cannot simultaneously measure all technological change and just the free gifts from externalities and scale effects.
- All improvements in technology, such as the internal combustion engine, do not "clearly raise TFP".
- TFP does not measure "prospects for longer term increases in output" since, among other reasons, (1) new GPTs tend to be associated with up-front costs and downstream benefits, and (2) there is no reason why the TFPs associated with successive GPTs should stand in any particular relation to each other.

[38] The standard AK models of economic growth require that formulation because it is the only one that allows positive R&D in a sustained or balance growth equilibrium (See Romer 1996, chapters three and four).

- There is reason to suspect that TFP does not adequately reflect the increase in a firm's capital value created by R&D activities that are realised through sale of intellectual property rather than exploitation by the developing firm. Yet these are often technological advances created by the use of valuable resources.
- TFP does not adequately capture the effects on the growth process of those technological changes that operate by lowering the cost of small industries and then allowing large subsequent increases in their sales and outputs.
- TFP does not adequately measure the massive amount of technological change that gets embodied in physical capital where under some circumstances the change is recorded as an increase in capital rather than as an increase in productivity.
- When full equilibrium does not pertain, as in the midst of any lagged adjustment process, the marginal equivalencies needed for successful aggregation do not obtain and it is likely that some of the increases in productivity of labour and capital will be recorded as increases in the quantities of labour and capital inputs.
- New technologies often lead to large up front costs of R&D and learning by doing and using that are incurred in the expectation of future benefits that will be missed when current outputs are related to current costs. The amount of this activity may vary with the life cycle of GPTs and other major technologies and, these variations will affect measured TFP.
- Neither TFP nor externalities adequately measure the technological complementarities by which an innovation in one sector confers benefit on other sectors—benefit for which those in other sectors would be willing to pay but do not have to do so.
- Low TFP numbers for the Asian Tigers do not mean they are in the same boat as was communist Russia; they are quite compatible with successful technology enhancing policies and technological transformation of a country through domestically generated or imported capital.
- TFP is as much a measure of our ignorance as it is a measure of anything positive.

It seems to us that, whatever TFP does measure—and there is cause for concern as to how to answer that question—it emphatically does not measure all of technological change. In the long term, we are interested in increases in output per unit of labour, resources (and waiting in the Austrian sense of the term). While people are of course free to measure anything that seems interesting to them, the degree of confusion surrounding TFP, particularly the assumption that low TFP numbers imply a low degree of technological dynamism, would seem to us to justify dropping the measure completely from all discussions of long term economic growth. Even if that does not happen, as we are sure it will not, every TFP measure should carry the caveat: *there is no reason to believe that changes in TFP in any way measure technological change.*

END OF TEXT

APPENDIX A: TYPES OF TECHNOLOGICAL COMPLEMENTARITY

The importance of technological complementarities, and the need to distinguish these clearly from externalities, justifies an extended (but still brief) consideration of the several types of complementarity.

Type 1

Specific technologies invented to solve problems in individual industries are often found to be applicable in a wide range of other industries.

Example 1 Machine tools[39]: 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 designed 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.

Type 2

The choice of a specific technique can give rise to a much enlarged space for radical and incremental innovation compared to the space made available to those choosing alternative methods of production. [40]

Example 2, The American System of Manufactures[41]: 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. This was resource intensive and labour saving and was based on standardized goods produced by specialized 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 could not have been foreseen, that the US system was capable of much more technological improvement than

[39] This example is based on Rosenberg (1976)

[40] 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.

[41] This example is based on "Why in America?" in Rosenberg (1994)

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 technology, the American system, enabled a much wider and deeper set of further inventions and innovations than did the choice of the alternative, the craft system.

Type 3

A single specific new machine can sometimes allow another whole industry to grow to mammoth size and, in its turn give rise to a host of other innovations and even other new industries, and also to a new form of organizing production.

Example 3, Automobiles[42]: One extreme example of technological complementarily is the result of the invention, early in the 20th century, of machine tools that could cut hardened steel. Previously machine tools could only cut soft metals. These were then hardened after manufacture, a process which unfortunately warped the parts which then had to be filed individually to make them fit. Henry Ford was the first car producer to see the potential of the new tools. He 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 got production of a standardized product based on interchangeable parts, it was a small step to introduce the assembly line, which created genuine mass production. (The assembly line could not have been introduced in a steam plant where machinery had to be arranged in the order of 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.) Ford's methods spread to virtually all other assembly operations in US manufacturing.

So there were two great technological complementarities in the introduction of the Fordist system. Machine tool developments, which were motivated by the needs of other industries, allowed for the production of standardised parts; electricity allowed for the assembly line. Incidentally, 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.

Type 4

Major GPTs that start with limited uses eventually spread to influence the entire economy, creating the possibility of countless new products, new process and new forms of organizing production and distribution.

[42] This example is based on Womack, Jones and Roos (1990).

Example 4, Electricity[43]: Electricity is a technology that made possible a vast range of products, processes and organizations that permeate today's entire production system and that is still giving rise to myriad innovations that could never have occurred in a non-electronic world of strictly mechanical technologies. In such a mechanical world, steam would have been the dominant source of power for decades after the time at which it actually gave way to electricity. The internal comb ustion engine would have become more and more important, as would engines powered by coal gas and natural gas. The electrification of factories would never have come. 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 computers of all sorts would have been un-thought of and mechanical calculators would have been all that was available for complex calculators—and no matter how sophisticated these are made, nothing mechanical can come within several orders of magnitude of matching the speed of electrons. Many of the household gadgets that revolutionized household management in the early 20th century, and displaced the host of servants formerly needed to staff a middle class household, would not have existed, although, some of these could have been powered by gas motors. Neither would we have electric lights, electric power, subways trains, and the host of other things that come to a halt every time power fails in a city. It is clear that electricity created an enormous set of technological complementarities in the sense that it created valuable opportunities for new innovation and investments in virtually all of the nation's industries, old and new.

Type 5

Some of the new technologies made possible by one GPT may evolve into GPTs in their own right.

Example 5, Electronic computers: Although enabled by electricity and in that sense subsidiary to it, electronic computers are on a par with electricity in the widespread revolution that they have enabled in almost all economic relations. Because we have elsewhere discussed the new innovation possibilities created by the computer, we only mention a few illustrative examples here.[44] Consider the robotization and automation of manufacturing, the restructuring of the firm from a hierarchical-pyramidal structure to a loose grouping of horizontally linked bodies, the Internet (itself prope rly regarded as a GPT in the extent that it permeates economy and creates new opportunities for innovation) which is part of the computer-based ICT revolution, modern methods of data retrieval and analysis, globalization of finance and production, the music industry, and the modern software industry. These are just a few areas where the computer has created a massive set of new possibilities for innovation—possibilities that simply would not have existed and would have been mostly un-thought of if the world had remained the mechanical world of the Victorian era.

[43] This example is mainly based on Nye (1990) and Schurr (1990).

[44] See e.g., Lipsey and Bekar (1995) and Lipsey (1999)

END OF APPENDIX

FIGURES

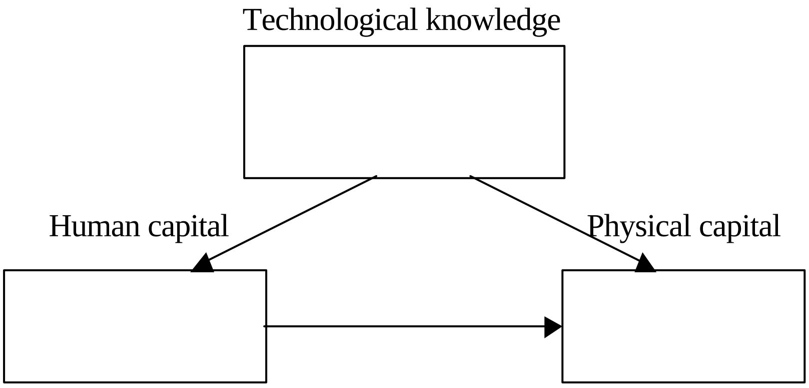

FIGURE 1

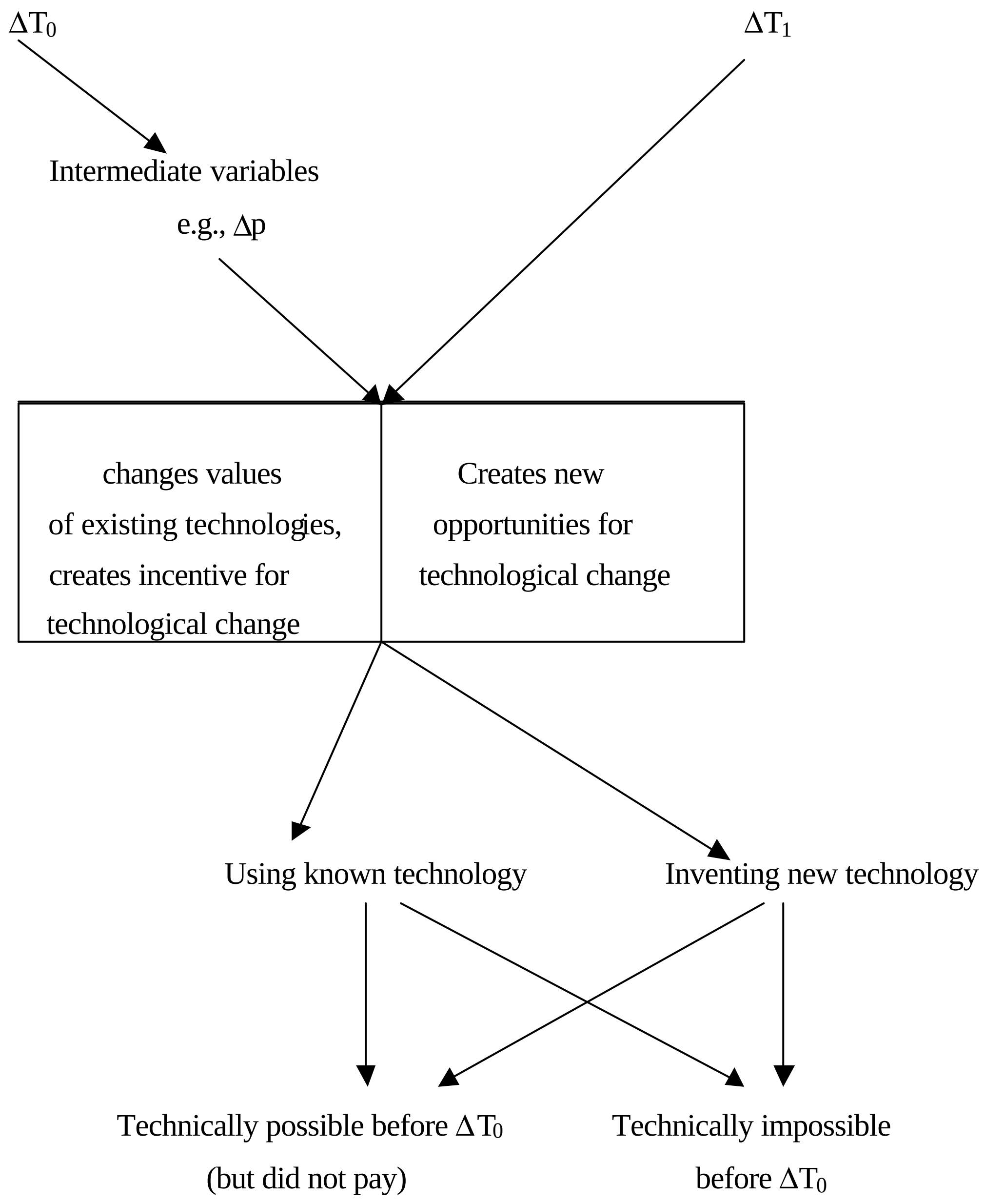
ΔT0
ΔT1
Intermediate variables
e.g., Δp
changes values
of existing technologies,
creates incentive for
technological change
Creates new
opportunities for
technological change
Using known technology
Inventing new technology
Technically possible before ΔT0
(but did not pay)
Technically impossible
before ΔT0

FIGURE 2

BIBLIOGRAPHY

David, Paul (1991), “Computer and Dynamo: The Modern Productivity Paradox in a Not too Distant Mirror”, in *Technology and Productivity: The Challenge for Economic Policy*, (Paris: OECD).

Eatwell, J., M. Milgate, and P. Newman (eds.) (1987), *The New Palgrave, a Dictionary of Economics*, (London: Macmillan).

Dollar and Wolf (1993), *Competitiveness, Convergence and International Specialization*, (Cambridge Mass.: MIT Press).

Griliches, S. (1994), “Productivity, R&D, and the Data Constraint”, *AER*, 84(1).

Griliches, S. (1995), “The Discovery of the Residual”. NBER Working Paper #5348

Grubler, Arnulf, (1998), *Technology and Global Change*, (Cambridge: Cambridge University Press).

Helpman, Elhanan (ed) (1998), *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press).

Jorgenson (1995) *Productivity, Volume 1: Postwar U.S. Economic Growth* (Cambridge: MIT Press)

Krugman, Paul, (1996) “The Myth of Asia’s Miracle” in *Pop Internationalism* (Cambridge: MIT Press)

Law, M. T. (2000), Productivity and Economic Performance: An Overview of the Issues, *Public Policy Sources*, No. 37.

Lipsey, Richard G., (1992) "Global Change and Economic Policy", in *The Culture and Power of Knowledge: Inquiries into Contemporary Societies*, Nico Stehr and Richard V. Ericson, (eds.) (Berlin: Walter de Gruyter).

_______________(1993) "Globalisation, Technological Change and Economic Growth", *Annual Sir Charles Carter Lecture* (Ireland: Northern Ireland Economic Council) Report #103.

_______________ (1999) “Sources of Continued Long-Run Economic Dynamism in the 21st Century” Chapter 2 in *The Future of the Global economy: Towards a Long Boom?* (Paris: OECD)

________________(2001) Success and Failure in the Transformation of Economics, *The Journal of Economic Methodology,* forthcoming

______________ and Clifford 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, (Kingston: John Deutsch Institute).

_______________ and Clifford Bekar and Kenneth Carlaw, (1998a) "What Requires Explanation", Chapter 2

____(1988b) “The Consequences of Changes in GPTs”, Chapter 7 in Helpman (1998).

Nye, D. (1990), *Electrifying America: Social Meanings of a New Technology, 1880-1940*, (Cambridge: MIT Press)

Pomeranz, Kenneth (2000*), The Great Divergence: China, Europe and the Making of the Modern World Economy*, (Princeton: Princeton University Press).

Romer, David (1996), *Advanced Macroeconomics*, (New York: McGraw Hill)

Rosenberg, Nathan (1976) “Technological Change in the Machine Tool Industry, 1840-1910”, in N. Rosenberg, *Perspectives on Technology*, (New York: M.E. Sharpe)

________________ (1982), *Inside the Black Box: Technology and Economics,* (Cambridge: Cambridge University Press).

________________(1994), *Exploring the Black Box: Technology, Economics and History,* (Cambridge: Cambridge University Press).

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

Statscan 13-568

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

Young, Alwyn (1992) “*A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore*” NBER Macroeconomic Annual (Cambridge: MIT press)

END OF MANUSCRIPT