Accounting for Trends in Productivity and R&D: A Schumpeterian Critique of Semi-Endogenous Growth Theory*

Joonkyung Ha — Korea Institute of Finance

Peter Howitt — Brown University

January 24, 2006

Abstract

This paper argues that long-run trends in R&D and TFP are more supportive of fully endogenous "Schumpeterian" growth theory than they are of semi-endogenous growth theory. The distinctive prediction of semi-endogenous theory is that sustained TFP growth requires sustained growth of R&D input. But in fact, TFP growth has been stationary since the early 1950s even though the growth rate of R&D input has fallen three-fold. Related to this, we show that the cointegration relations implied by semi-endogenous theory are not found in the data and that the theory has consistently predicted a fall in trend TFP growth that has not materialized. In contrast, the prediction of Schumpeterian theory that sustained TFP growth requires a sustained fraction of GDP to be spent on R&D is not contradicted by experience, the cointegration relations implied by Schumpeterian theory are found in the data, and although Schumpeterian theory does not fit the observed time path of TFP much better than semi-endogenous theory it outperforms semi-endogenous theory in predicting TFP growth out of sample.

*Chris Laincz and Marios Zachariadis provided helpful comments and suggestions. We also thank Chad Jones for his constructive criticisms.

1 Introduction

The main empirical basis of the "R&D-based" branch of endogenous growth theory is the well-established finding that TFP growth is strongly influenced by R&D expenditures.[1] Yet one of the greatest challenges facing the theory has been how to account for long-run co-movements in R&D and TFP within OECD countries. In particular, Jones (1995a,b) has observed that since the early 1950's the number of scientists and engineers engaged in R&D in the United States has grown more than fivefold without resulting in any increase in the growth rate of output per person or of TFP. The coexistence of an upward trend in R&D labor and no trend in TFP growth has effectively refuted the first generation of R&D-based theories,[2] according to which more R&D labor should induce more TFP growth.

Theorists have responded to the challenge by developing a second generation of R&D-based theories that are consistent with these findings. The new theory comes in two varieties, with similar foundations but radically different long-run implications. One is the "semi-endogenous" theory of Jones (1995b), Kortum (1997) and Segerstrom (1998), which modifies the original theory by incorporating diminishing returns to the stock of knowledge in R&D. That is, as technology develops and becomes increasingly complex, sustained *growth* in R&D labor[3] becomes necessary just to maintain a given rate of TFP growth. Semi-endogenous growth theory has a stark long-run prediction, namely that the long-run rate of TFP growth, and hence the long-run growth rate of per-capita income, depends on the rate of population growth, which ultimately limits the growth rate of R&D labor, to the exclusion of all economic determinants.

The other variety of second-generation R&D-based theory consists of the fully-endogenous "Schumpeterian" models of Aghion and Howitt (1998a, ch.12), Dinopoulos and Thompson (1998), Peretto (1998) and Howitt (1999). These authors all build on Young's (1998) insight that as an economy grows, proliferation of product varieties reduces the effectiveness of R&D aimed at quality improvement, by causing it to be spread more thinly over a larger number of different sectors. Without denying the deleterious effect of increasing complexity on the productivity of R&D,[4] Schumpeterian theory retains the original assumption of constant returns to the stock of knowledge in R&D, and therefore implies that the long-run rate of TFP growth will be governed by the same economic factors as in the first-generation R&D-based theories, with the sole exception that the size of a country's labor force no longer has a positive scale effect on long-run growth. The theory is consistent with the observed coexistence of stationary TFP growth and growing R&D

[1] See for example Griliches (1994) and Zachariadis (2003).

[2] Segerstrom, Anant, and Dinopoulos (1990), Romer (1990), Corriveau (1991), Grossman and Helpman (1991), and Aghion and Howitt (1992).

[3] Or human capital, which leads to the same long-run implications.

[4] More specifically, Aghion and Howitt add two considerations that offset increasing complexity: (a) because capital is also an important input to R&D therefore capital deepening induced by productivity growth helps to offset diminishing returns in R&D, and (b) labor-augmenting technological progress that makes workers more productive in manufacturing also tends to make them more productive in R&D. These points will be elaborated on in more detail in Section 2.1 below.

labor, because growing R&D labor is needed to counteract the deleterious effect of product proliferation on the productivity of R&D.

The purpose of this paper is to compare these two varieties of second-generation R&D-based growth theory, by asking which of them is more consistent with observed TFP trends during the second half of the 20th Century. Our focus will be on the US experience, because that is the case where we have the longest time-series on R&D. At stake is the central idea of endogenous growth theory, namely that long-run economic growth has economic determinants, which in R&D-based theories consist of the policies, regulations and institutions that impinge on the incentives to create and adopt new technologies. Schumpeterian theory confirms this idea and semi-endogenous theory denies it.[5]

Our analysis also bears on the debate over how to view the productivity acceleration of the 1990s and the "new economy." Some (for example, Maddison, 2001 and Gordon, 2002) have argued that the new economy was at most a temporary boom due to transitory factors, while others (Baily and Lawrence, 2001; Easterly and Levine, 2001; Basu, Fernald and Shapiro, 2001) have argued that the acceleration of productivity appears to have been more than temporary. Semi-endogenous growth theory predicts that the pessimistic forecast will ultimately prevail because the world's population cannot continue to grow without bound.[6] On the other hand, Schumpeterian theory predicts that there will be no slowdown in the rate of TFP growth in the absence of changes in the economic determinants of R&D intensity, regardless of what happens to global population growth.

Our main finding is that Schumpeterian theory is more consistent with the long-run trends in R&D and TFP than semi-endogenous theory. Specifically, since 1953 the growth rate of R&D labor, as measured by the number of scientists and engineers engaged in R&D in G5 countries, or by the number engaged in R&D in the United States, has fallen more than threefold without resulting in any dramatic reduction in the growth rate of TFP in the United States. This fact undermines the central proposition of semi-endogenous growth theory, for if sustained growth in TFP does not require sustained growth in R&D labor then there is no reason to believe that long-run TFP growth is governed exclusively by population growth.

Although the downward trend in the growth rate of R&D labor casts doubt on semi-endogenous growth theory, it does not contradict Schumpeterian theory, according to which trendless TFP growth requires only the absence of a trend in the proportion of GDP spent on R&D. Indeed there is no such trend; the R&D/GDP

[5] Our paper contributes to a larger ongoing debate regarding the empirical relevance of Schumpeterian theory. See Pack (1994), Coe and Helpman (1995), Jones (1995a, 1995b), Aghion and Howitt (1998b), Parente and Prescott (1999), Howitt (2000), Arora (2001), Zachariadis (2003), Aghion et al. (2005), Aghion, Howitt and Mayer-Foulkes (2005) and Howitt and Mayer-Foulkes (2005). The paper by Zachariadis is particularly noteworthy in this regard, as it finds direct evidence that R&D expenditures as a fraction of GDP, one variant of the product-proliferation-adjusted R&D input that should drive productivity growth according to Schumpeterian theory, is indeed what drives productivity growth in a panel of U.S industries.

[6] Indeed the specific semi-endogenous model of Jones (2002) implies that, even without a slowing of world population growth, about 80% of the US economic growth in the second half of the 20th Century was due to temporary factors such as increases in educational attainment and research intensity.

ratio in the United States has remained between 0.021 and 0.029 every year from 1957 through 2000.

We conduct two other related tests of the competing theories, both arriving at the same conclusion. The first consists of cointegration tests. Under semi-endogenous theory, the observed stationarity of productivity growth implies that the logs of productivity and R&D input should be cointegrated. Under Schumpeterian theory it implies that the logs of GDP and R&D should be cointegrated with a coefficient of unity. We find strong evidence of the latter but none of the former. The second test consists of seeing which theory forecasts productivity better out of sample. We find that semi-endogenous theory does much worse, because it always predicts a fall in productivity growth that has not materialized; on the contrary, productivity growth had an upward trend over the final quarter of the 20th Century.

Two other studies have tested Schumpeterian theory with similar results using US data. Zachariadis (2003) finds that patenting and productivity growth are positively affected by the share of output devoted to R&D in a panel of US industries over the period 1963-88. This adjusted measure of R&D is the industry-level counterpart to the adjusted measure which Schumpeterian theory predicts should matter for aggregate productivity growth, namely the R&D to GDP ratio.[7] Laincz and Peretto (2004) note that according to Schumpeterian theory if the economy grows at a steady rate there should be no increase in the amount of R&D per product line, and that fluctuations in output should be related to fluctuations in employment per product line and in R&D per product line; they present evidence corroborating these predictions using R&D personnel per establishment as a proxy for R&D per product line over the period 1964-2001. Another relevant study is that of Ulku (2005), which estimates the degree of returns to the stock of knowledge in R&D in an R&D-based growth model using panel data from 41 countries over the 1981-97 period, and finds that it is almost exactly equal to unity, the value predicted by Schumpeterian theory.

Section 2 below presents a summary sketch of semi-endogenous and Schumpeterian theory, deriving their respective long-run predictions concerning R&D and TFP growth. Section 3 describes the relevant facts. Section 4 subjects the two theories to formal cointegration tests. Section 5 asks which theory forecasts productivity better out of sample. Section 6 discusses the idea of omitting the space and defense components of R&D when testing the theories. Section 7 tests an augmented version of the semi-endogenous model that incorporates the product proliferation effect, and discusses the notion of "double knife-edge." Section 8 offers some concluding remarks.

[7] Zachariadis (2004) finds that the same Schumpeterian model is also supported by aggregate data from a panel of 10 OECD countries.

2 Different versions of R&D-based growth theory

The central component of any R&D-based growth model is a knowledge-creation function, according to which the flow $\dot{A}$ of new knowledge depends on R&D input X and other variables.[8] In these models, A measures productivity[9] as well as knowledge, so that specifying a knowledge-creation function is equivalent to specifying a productivity-growth function, according to which the growth rate $g_A \equiv \dot{A}/A$ of TFP is a function of R&D input and other variables. The key differences between competing models can be interpreted formally as different assumptions on these two related functions, which are typically specified as Cobb-Douglas.

2.1 First-generation fully-endogenous models

In the models developed by Aghion and Howitt (1992), Grossman and Helpman (1991) and Romer (1990) the productivity-growth function is:

$$g_A = \lambda X^\sigma \tag{1}$$

where $0 < \sigma \leq 1$. Equivalently, the knowledge-creation function is:

$$\dot{A} = \lambda X^\sigma A \tag{2}$$

R&D input X is specified as either the flow N of labor allocated to R&D, or the productivity-adjusted flow R/A of R&D expenditure,[10] on capital as well as labor. The models all assume a constant population size and possess a balanced-growth equilibrium in which X is constant. Long-run growth thus depends on policies and other parameters that determine the long-run level of R&D input.

The unit exponent on A in the right hand side of (2) imposes constant returns to knowledge in the production of new knowledge. In the quality-improvement version of Aghion and Howitt (1992) this follows from the assumption that each innovation produces a fixed proportional quality improvement.[11] Constant returns is what allows sustained, non-explosive endogenous growth.

When R&D input is specified as productivity-adjusted R&D expenditure (R/A), innovations are typically assumed to be produced according to the same production function as consumption and capital goods. The division of R by A recognizes the force of increasing complexity. That is, as technology advances it takes an ever-increasing R&D expenditure just to keep innovating at the same rate. Constant returns to knowledge

[8] In theory, the input to knowledge production should include all knowledge-creation activities broadly defined, not just R&D. In practice however, R&D data provide the broadest available measure of input to knowledge creation.

[9] For brevity, but in violation of convention, we define productivity in labor-augmenting terms and refer to it as TFP.

[10] See Rivera-Batiz and Romer (1991), Barro and Sala-i-Martin (1995, ch.7) or Howitt and Aghion (1998). R/A is comparable with N in the sense that along a balanced growth path both variables will grow in proportion to population.

[11] In the product-variety model of Romer (1990), constant returns amounts to a special assumption on the knowledge spillover whereby an increase in the number of extant varieties facilitates the generation of new varieties. For concreteness we use the Aghion-Howitt quality-improvement interpretation in what follows.

prevail in the long-run because along a balanced growth path, with R/A constant, as A rises the negative effect on productivity growth due to increasing complexity is just offset by the positive effects due to capital deepening and increased labor-efficiency, both of which raise productivity in R&D as well as in manufacturing (that is, both allow R&D expenditure R to rise relative to $R\&D$ labor).

The growth equation (1) implies that in a setting where the long-run growth rate of R&D input is positive, the long-run growth rate of productivity should be rising. This is why the observation that R&D labor has a positive trend growth rate while TFP growth is stationary refutes first-generation models. Defining X as productivity-adjusted R&D expenditure does not rescue the models because this definition of R&D input also has a positive trend growth rate over the second half of the 20th Century.[12]

2.2 Semi-endogenous models

To avoid this problem, Jones (1995b) developed an elegant semi-endogenous model with decreasing returns to knowledge:

$$g_A = \lambda X^{\sigma} A^{\phi - 1}, \qquad \phi < 1 \tag{3}$$

where again X is R&D input. Taking logs and differentiating both sides of (3) with respect to time yields:

$$\frac{\dot{g}_A}{g_A} = (1 - \phi)\left(\gamma g_X - g_A\right) \tag{4}$$

where $\gamma \equiv \frac{\sigma}{1-\phi}$ and g_X is the exponential growth rate of R&D input.

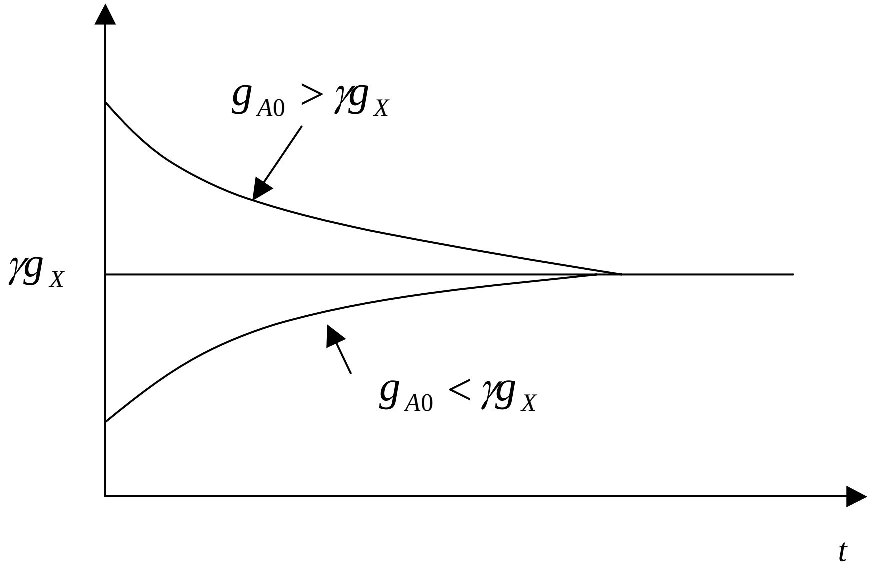
This semi-endogenous model is compatible with the observation of positive trend growth in R&D input, because as long as $\phi < 1$ and the time path of g_X is bounded, the differential equation (4) yields a bounded solution for TFP growth. In particular, if g_X is constant, or approaches a constant, then:

$$g_A(t) \rightarrow \gamma g_X.$$

In the long run the growth rate of R&D labor cannot exceed the growth rate n of population, and in a balanced growth equilibrium it will equal n. Likewise, the growth rate of productivity-adjusted R&D expenditure will equal n along a balanced growth path. This leads to the radical implication that the long-run growth rate of an economy will equal γn, irrespective of what fraction of society's resources is assigned to knowledge creation. Policies to stimulate R&D will have at most transitory effects on productivity growth and, by extension, on growth of per-capita income.

[12] See Figure 5 below.

Figure 1: Time path of g_A when γg_X is constant



2.2.1 Other dynamic implications

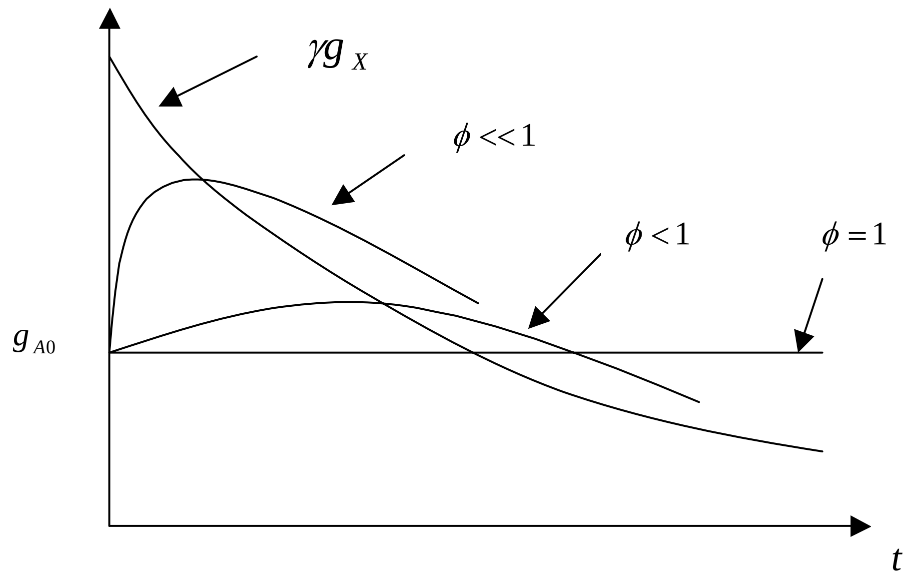
Even before we approach the long run in which growth of R&D input is limited by population growth, semi-endogenous theory has testable implications regarding the relationship between productivity growth g_A and the growth rate g_X of R&D input. First, when g_X is constant, productivity growth will be monotonically decreasing if it is initially above the quasi-equilibrium value γg_X, and monotonically increasing if it is initially below γg_X, as shown in Figure 1. The speed with which g_A converges to γg_X depends inversely on the degree of returns to scale ϕ, and approaches zero as ϕ approaches the limiting value of unity that defines the special case of first-generation theory.

Second, when the growth rate of R&D input is falling over time, productivity growth will be monotonically decreasing if it begins above γg_X, and will follow an inverse-U time path if it begins below γg_X, as shown in Figure 2. In the latter case the steepness of the inverse-U time path of productivity will depend negatively on the degree of returns to scale ϕ. In the limiting special case of first-generation theory ($\phi = 1$) the inverse-U becomes a horizontal line, with constant growth and no convergence.[13] If ϕ is far from unity then the path of productivity growth will be almost proportional to the path of the growth rate of R&D input.

In summary, semi-endogenous theory implies that sustained productivity growth requires sustained growth in R&D input. If, on the contrary, the growth rate of R&D input is falling then the theory predicts that productivity growth will follow an inverse-U shaped time path, eventually falling, except in the limiting

[13] Note that if we hold γ constant in this thought experiment, then as ϕ approaches unity the R&D elasticity $\sigma = (1 - \phi)\gamma$ approaches zero, so the model is approaching the neoclassical limit of exogenous productivity growth: $g_A = \lambda$.

Figure 2: Time path of g_A when γg_X is decreasing



case that defines the already refuted first-generation of R&D-based models. This implication is critical to semi-endogenous theory, because if productivity growth could be sustained even when the growth rate of R&D labor is falling, then there would be no reason to believe the theory's main distinguishing proposition, namely that population growth ultimately governs productivity growth.

2.3 Fully-endogenous models with product proliferation

The other version of R&D-based theory that is compatible with growing population and growing R&D labor is the Schumpeterian version exemplified by the models presented in Aghion and Howitt (1998a, ch. 12) and Howitt (1999). These models retain the assumption of constant returns to knowledge but take into account the deleterious effect on productivity growth arising from product proliferation, as pointed out by Young (1998), by specifying $g_A = \lambda\,(X/Q)^{\sigma}$, where Q is a measure of product variety. In these models Q is typically proportional to the size L of population, so we have:

$$g_A = \lambda\,(X/Q)^{\sigma}\text{ , or } g_A = \lambda\,(X/L)^{\sigma} \tag{5}$$

An economic interpretation of (5) is that a larger population increases the number of people who can enter an industry with a new product, thus resulting in more horizontal innovations, which dilutes R&D expenditure over a larger number of separate projects.

This theory is consistent with the observed coexistence of stationary TFP growth and rising R&D input, which refuted the first generation theory, because in the long-run X and L grow at the same rate. Thus the growth-enhancing effect of rising R&D input is just offset by the deleterious effect of product proliferation.

More generally, Q can depend on any variable that grows in the long run at the same rate as population – for example, hL where h is human capital per person, or productivity-adjusted output Y/A which is a (productivity-adjusted) measure of aggregate purchasing power. Thus we can have X/Q equal to N/L, N/hL, R/AL, R/AhL or R/Y. In the last of these cases the propensity to enter with a new product would depend on the productivity-adjusted level of GDP, as in a model of external learning-by-doing (with an adjustment to account for the effect of increasing complexity).

Because (5) retains the assumption of constant returns to knowledge, this theory preserves most of the other long-run implications of first-generation theory. In particular, it implies that anything that raises the fraction of society's resources allocated to R&D will raise the long-run TFP growth rate. Accordingly we refer to the class of models incorporating product proliferation and constant returns to knowledge as second-generation fully-endogenous theory or "Schumpeterian" theory.

Schumpeterian theory differs from semi-endogenous theory in terms of the necessary conditions it implies for sustaining long-run productivity growth. According to semi-endogenous theory, as we have seen, society must sustain growth in R&D input. But according to Schumpeterian theory society must just sustain the fraction of resources it allocates to R&D. For example, if X/Q is interpreted as N/L then (5) implies that productivity growth will remain constant provided that the fraction of labor allocated to R&D remains constant. Likewise, if X/Q is interpreted as R/Y then productivity growth will remain constant provided that the fraction of GDP spent on R&D remains constant.

2.4 Summary and remarks on the source of growth

All these models can be interpreted as special cases of a general model in which productivity growth obeys:

$$\begin{aligned} g_A &= \lambda \left(X/Q\right)^{\sigma} A^{\phi-1} \qquad (6) \\ Q &\propto L^{\beta} \text{ in the steady state} \end{aligned}$$

They differ only in the restrictions they place on the parameters of (6), most importantly the degree of returns to scale in knowledge ϕ and the coefficient of product proliferation β. Table 1 summarizes the differences.

Table 1. Classifying various growth models

	ϕ	β	σ
Neoclassical	$= 1$	–	$= 0$
Fully-endogenous I	$= 1$	$= 0$	> 0
Semi-endogenous	< 1	$= 0$	> 0
(with product proliferation)	(< 1)	(< 1)	(> 0)
Fully-endogenous II	$= 1$	$= 1$	> 0

The first line shows that even the neoclassical model can be considered a limiting special case of (6). The difference between it and the other (endogenous) models is that it specifies a zero coefficient (σ) on R&D input, so that technological progress is exogenously given. The value of this coefficient is not important, however, for distinguishing between endogenous models.

Important distinguishing features of the different endogenous theories have to do with (a) the degree of returns to scale in knowledge production ϕ, and (b) the coefficient of product-proliferation β. Semi-endogenous theory is characterized by diminishing returns to knowledge and no (or imperfect) product-proliferation effect, whereas (second-generation) fully-endogenous theory builds on the assumption of constant returns to knowledge and a complete product-proliferation effect.[14]

In order to see the different implications on the ultimate source of growth, consider the special case where $X = N = vL$ and $Q = L^{\beta}$, and where the fraction v of labor allocated to R&D is assumed to be constant in the steady state. Then, equation (6) can be rewritten as:

$$g_A = \lambda \left(vL^{1-\beta} \right)^{\sigma} A^{\phi - 1}$$

from which we can derive simple versions of the three endogenous growth models as follows.

$$\text{Fully-endogenous I } (\phi = 1 \text{ and } \beta = 0) : g_A = \lambda \left(vL \right)^{\sigma} \tag{7}$$

$$\text{Semi-endogenous } (\phi < 1 \text{ and } \beta = 0) : g_A = \lambda \left(vL \right)^{\sigma} A^{\phi - 1} \rightarrow \frac{\sigma}{1 - \phi} n \tag{8}$$

$$\text{Fully-endogenous II } (\phi = 1 \text{ and } \beta = 1) : g_A = \lambda v^{\sigma} \tag{9}$$

The first generation fully endogenous model in equation (7) has growth depending always on the *level* of both v and L, while the semi-endogenous model in equation (8) has long-run growth depending on the *growth* of L as opposed to its level. In contrast, the second generation fully endogenous model in equation

[14] The case of semi-endogenous model with (imperfect) product proliferation is analyzed in Section 7 below.

(9) shows the effects of only v, the fraction of workers allocated to R&D.

3 Trends in productivity and R&D

This section examines the long-run trends in the data with a view to seeing which theory is more compatible with them. Figure 3 below shows the log of US TFP[15] over the half-century from 1950 to 2000, along with a smoothed trend constructed using the Hodrick-Prescott filter.[16] Figure 4 shows the corresponding growth rate of TFP. These figures reveal a clear upward trend in the level of productivity, but not in its growth rate. Instead, productivity growth tended to slow down over the 1964-1975 period and then to increase over the 1975-2000 period, with no overall trend.

As indicated in the first column of Table 2 below, the trend change in productivity growth was a statistically insignificant decrease of only .01 percentage points per year, which is just 3/4 of a percent of the sample average productivity growth rate. (Even this small trend change is attributable to the rapid burst of growth in the very early 1950s. If we take out the first data point (1951) the trend change becomes 0.00.)

Table 2. Trends in productivity growth and growth of R&D input

Variable (X)	g_A	g_N (G5)	g_N (US)	$g_{R/A}$
Average value[a]	1.40	4.73	4.20	3.69
Trend annual change[b]	-0.01	-0.14	-0.13	-0.15
p-value[c]	0.614	0.000	0.012	0.085
Trend annual change[a] as a percentage of the average value	-0.75	-2.87	-3.08	-4.13

A is US Productivity, N is the number of scientists and engineers engaged in R&D and R is US real R&D expenditure. The sample period is 1953-2000 for R/A and 1950-2000 for all other variables. See Appendix A for description of data.
[a]Expressed in percentage points
[b]OLS estimate of β in the regression: $X = \alpha + \beta \cdot year$
[c]Using HAC standard errors

3.1 Gross measures of R&D input

If the data were generated by a semi-endogenous growth model, then this constancy in productivity growth for half a century would imply that there should be no trend in the growth rate of R&D input. But Figures 5 and 6 show that this has not been the case. Figure 5 plots, in logs, three different measures of R&D input, namely productivity-adjusted US real R&D expenditures (R/A), the number of scientists and engineers

[15]See Appendix A for the details of this TFP calculation.
[16]The smoothness parameter (lambda) of the HP filter was set to 100.

Figure 3: Trend of the level of productivity (log), US

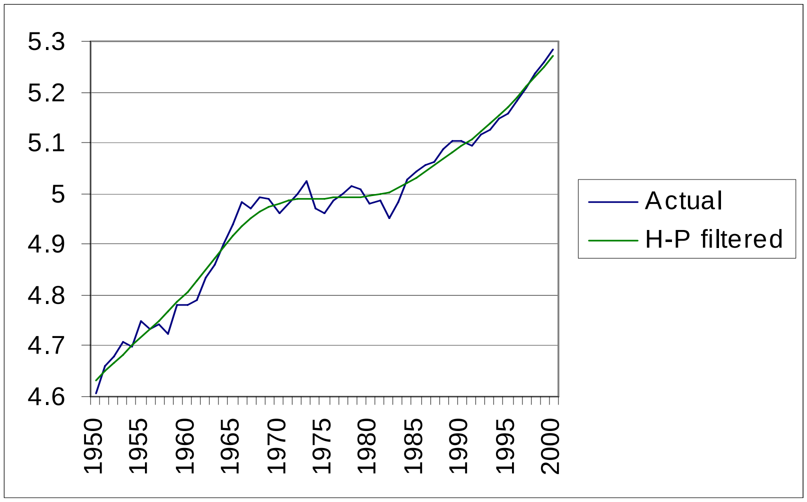
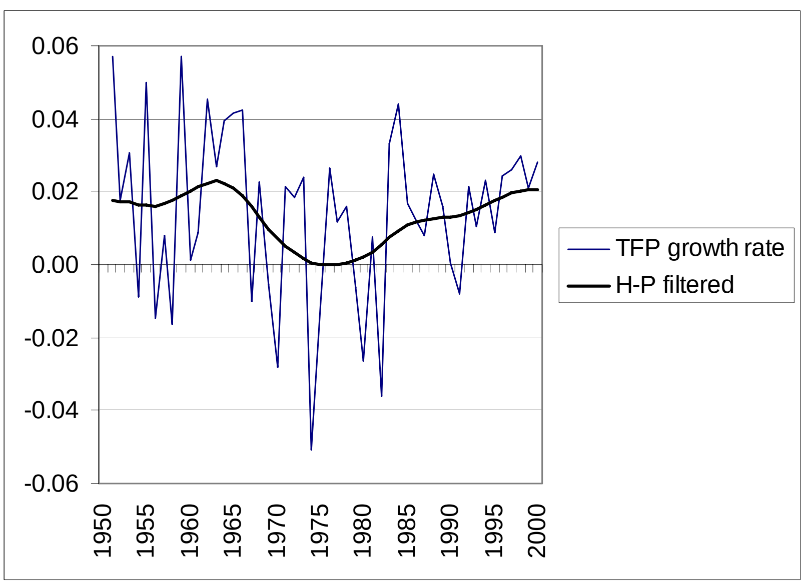
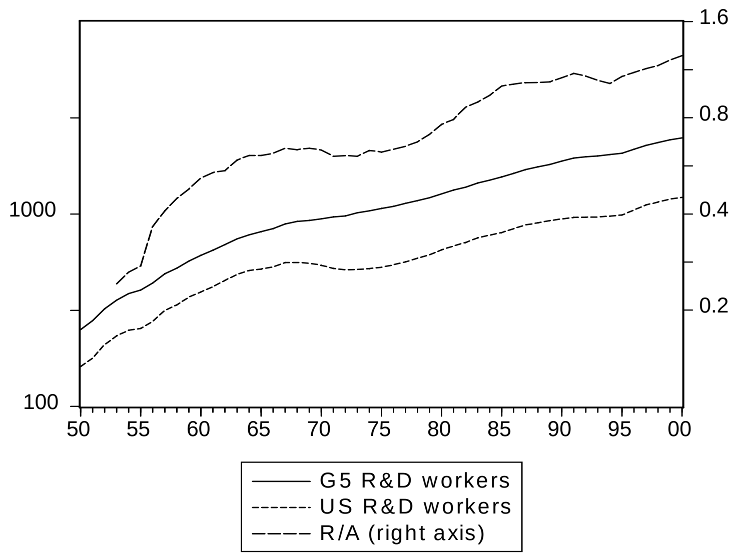


Figure 4: TFP growth rates, US, 1950-2000



engaged in R&D in the United States, and the number of scientists and engineers engaged in R&D in the G5 countries. The last of these is Jones's (2002) preferred measure, as it takes into account the likelihood that much of the relevant input to US productivity growth comes from ideas generated in other leading industrial countries.[17] Figure 6 plots the annual growth rates of these three measures. The figures show clear evidence of logistic growth in R&D input, with an upward trend in the level but a downward trend in the growth rate.

Figure 5: R&D workers and R&D expenditures (log) 1950-2000

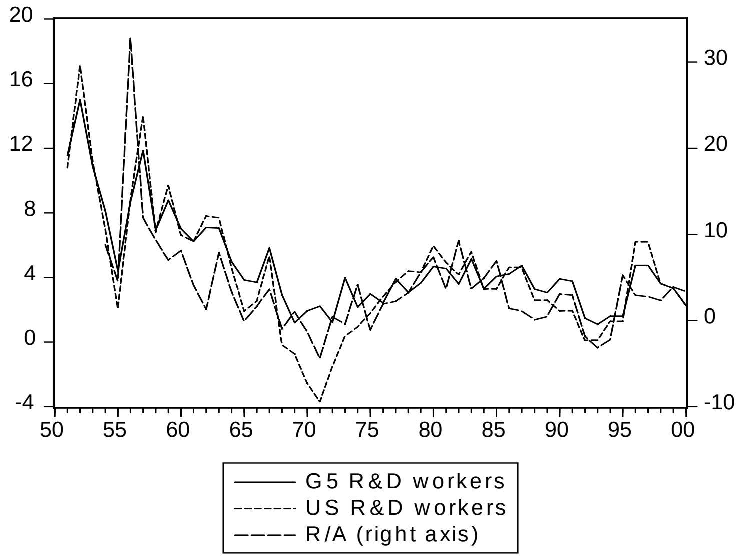


Columns 2 through 4 of Table 2 confirm this by showing that the least-squares linear trends imply a substantial annual decrease in the growth rate of each measure of R&D input. The estimated trend decrease in growth rate was between 2.87 and 4.13 percent of the corresponding average growth rate for each measure, and in all cases the estimated decrease was statistically significant.

The fact that the growth rate of R&D input has been decreasing substantially without any corresponding substantial decrease in the growth rate of productivity is not consistent with the predictions of semi-endogenous growth theory. As was explained in Section 2.2 and Figure 2, the fall in the growth rate of R&D input should have produced either a fall in TFP growth or at least an inverse-U shaped time path of TFP growth. Instead, the HP-smoothed time series of productivity growth displayed in Figure 4 shows a U-shaped path, with TFP growth falling before 1975 and then rising afterwards, and no overall trend.

[17] We also include the measure of R&D scientists and engineers in the United States because although it excludes inputs from abroad it is measured more acurately during the period from 1950 to 1965. Over this period there were no data for G5 countries other than the United States, so we used Jones's (2002) estimates for these years. The fact that N_{US} is such a large proportion of N_{G5} (64% in 1965, falling to 50% in 2000) implies that the two series behave quite similarly for our purposes.

Figure 6: R&D workers and R&D expenditures (growth rates) 1950-2000



One possible explanation for this failure of semi-endogenous theory is that the first half of the sample period was a period of transition, during which the long-term relationships implied by the theory were not observed because of the lags which are almost certainly present in the effect of R&D on productivity growth, and which we have not taken into account. However, even if we consider only the second half of the sample period the theory does not seem to fit the facts very well because productivity growth showed a trend *increase* over this 25 year period, even though there was apparently no substantial reversal of the downward trend in the growth rate of R&D input. As mentioned before, the theory predicts that productivity growth should have followed an inverse-U shaped path, with a trend *decrease* over the second half.

It is also possible that the entire second half of the 20th Century was a transitional period too short to reveal the long-run relationships predicted by semi-endogenous theory, But if this is the case then the lags are clearly so long that the theory's distinguishing long-run predictions have little policy relevance and are untestable with available data on productivity and R&D.

3.1.1 Behind the logistic growth of R&D input

The logistic growth in R&D input is apparently connected with the fact that total number of scientists and engineers has been leveling off in the United States. According to the NSF, the total number was 3.30 million in 1993, 3.19 million in 1995, and 3.37 million in 1997. This total is what places the upper limit for the number of scientists and engineers engaged in R&D in the biggest R&D-performing country, the United States.

There are several notable facts behind this total.[18] First, the absolute number of students in the field of engineering declined. During the last two decades of the 20th Century, the college-age population in the United States declined by more than 21 percent: from 21.6 million in 1980 to 17.0 million in 2000. Echoing this overall demographic decline, the number of students enrolling in undergraduate engineering decreased by 16%, from a high point of 441 thousand students in 1983 to 356 thousand in 1996.[19] Second, the only fields with an increasing number of earned degrees in the 1990s have been psychology and biological sciences. Entry into these fields has offset the overall demographic downturn. Engineering, mathematics, and computer science fields show declining numbers of degrees in the late 1980s and the 1990s. Third, as degree production grows at a slower rate than in the past, the average age of scientists and engineers in the labor force increased - with mixed implications for different aspects of research productivity.

Despite these facts, total number of R&D workers seems to have increased in the late 1990s, probably due to the substantial immigration of R&D workers from developing countries like China and India.[20] In the United States, there was a large immigration of professional specialists (138 thousand in 1996-97) and temporary foreign workers with H1B visas (385 thousand in 1996-98) in the 1990s. Those from developing countries, who did not show up in the earlier periods, have been added to the data in the 1990s.

Considering all these facts, it seems that the slowdown in growth of the number of research workers is not just a measurement or data problem and that without large increases in immigration it is likely to continue for decades to come.[21]

3.2 R&D measures adjusted for the product proliferation effect

If the data were generated by a Schumpeterian growth model, then constancy in productivity growth for half a century would imply that there should be no trend in the level of adjusted measures of R&D input; that is, measures that have been adjusted to take into account the effect of product proliferation. According to Figures 7 and 8, there does appear to be a slight upward trend in these adjusted R&D measures. However, as a comparison of the last rows of Tables 2 and 3 confirms, the trend increase was much smaller, relative to the average value of the corresponding measure, than the trend decrease in the growth rate of gross R&D input. Moreover, all the measures of adjusted R&D input that use R&D expenditure as the gross measure have an upward trend that is less than 0.8 percent of the measure's average value, and two of the three have statistically insignificant trends. Thus, in contrast to semi-endogenous theory, there are at least some measures of adjusted R&D input (specifically R/Y and R/AhL) whose trends clearly behave the way fully

[18] These facts can be found in the NSF website.

[19] Trends in graduate engineering enrollment differ a little. Graduate enrollment increased from 1979 to 1993, but has declined each year since.

[20] Data can be obtained in the INS website.

[21] See Appendix B for more details of the trends in the number of R&D workers.

endogenous theory says they must, given the observed trendless behavior of productivity growth.[22]

Table 3. Trends in productivity growth and product-proliferation-adjusted R&D input

Variable (X)	g_A	N/L (G5)	N/L (US)	N/hL (G5)	N/hL (US)	R/AL	R/AhL	R/Y
Average value[a]	1.40	1.23	0.67	0.55	0.30	7.96	3.55	2.43
Trend annual change[b]	-0.01	0.02	0.01	0.01	0.00	0.06	0.01	0.01
p-value[c]	0.614	0.000	0.000	0.000	0.003	0.004	0.447	0.160
Trend annual change[a] as percent of average value	-0.75	1.98	1.32	1.46	0.77	0.78	0.22	0.40

A is US Productivity, N is the number of scientists and engineers engaged in R&D, L is the number of employed US workers, h is human capital per US worker, R is US real R&D expenditure and Y is US real GDP. The sample period is 1953-2000 for variables involving R, 1950-2000 for all others. See Appendix A for description of data.

[a]Expressed in percentage points

[b]OLS estimate of β in the regression: $X = \alpha + \beta \cdot year$

[c]Using HAC standard errors

3.3 Summary

The fact that there was no substantial trend in US productivity growth over the second half of the 20th Century provides a simple test of both semi-endogenous growth theory and second-generation fully-endogenous growth (Schumpeterian) theory, and thus provides a simple way of comparing the empirical performance of the two competing theories. Given the absence of a trend in productivity growth, semi-endogenous theory implies that there should be no substantial trend in the growth rate of R&D input, whereas Schumpeterian theory implies that there should be no substantial trend in the level of product-proliferation-adjusted R&D input. In this section we showed that the data tend to bear out the implication of Schumpeterian theory, at least for some definitions of the variables, but they do not bear out the implications of semi-endogenous theory for any definition of the variables. For while adjusted R&D input had only a modest upward trend, the growth rates of all measures of gross R&D input had a large and statistically significant downward trend. Figure 9 gives an indication of the disparity between the two theories by putting on the same scale two variables which the respective theories imply should have been trendless, both normalized to have the same average value (.01396) as the growth rate of productivity. The small upward trend in the ratio of R&D expenditures to GDP is hardly visible against the backdrop of the large downward trend in the growth rate of the number of scientists and engineers engaged in R&D.

[22]Of course it would be possible by dropping selected categories of spending from our measure of aggregate R&D to construct a narrower measure of adjusted input with a substantial trend. In particular, removing the defense and space related component, which grew most rapidly in the late 1950s and into the 1960s, would do the job. We discuss this possibility in section 6 below.

Figure 7: Product-proliferation-adjusted R&D workers (log)

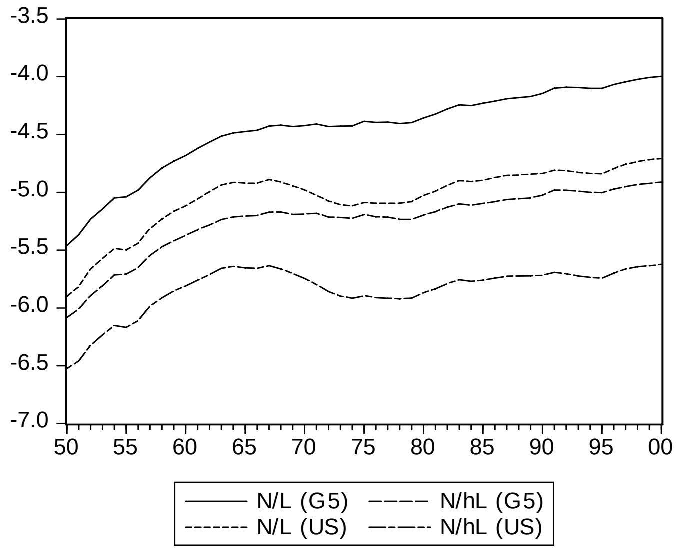


Figure 8: Product-proliferation-adjusted R&D expenditures (log)

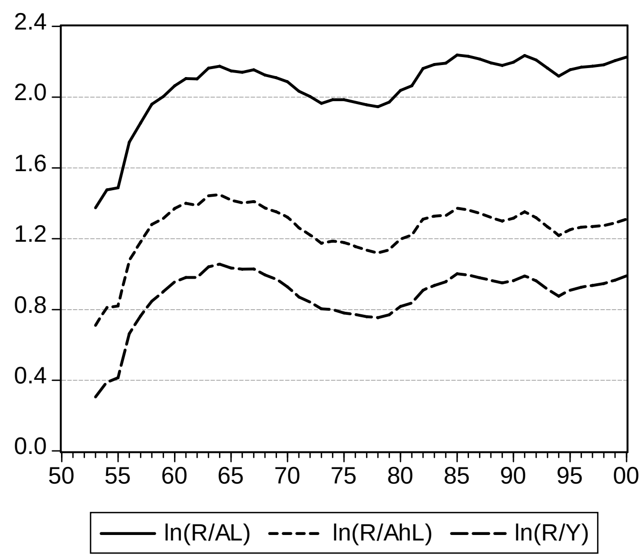
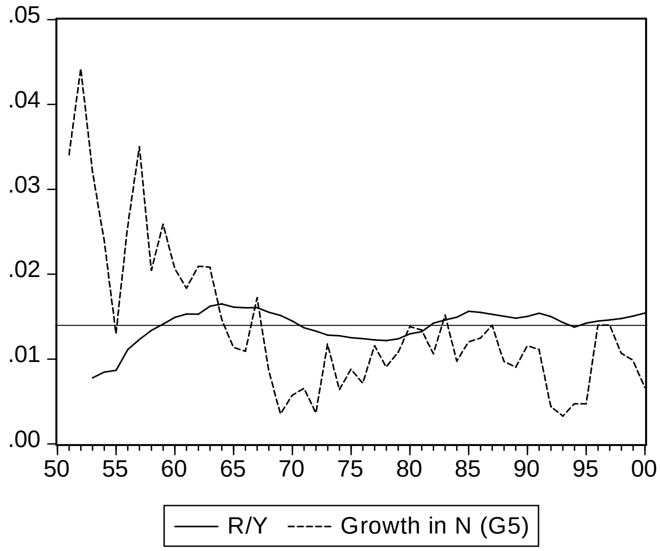


Figure 9: R&D to GDP ratio and growth in G5 R&D workers



4 Cointegration Tests

According to each of the two competing theories, the stationarity of productivity growth implies specific cointegrating relationships that should exist in the data. Thus another way to compare the theories is to see which of these cointegrating relationships can be detected using standard time-series tests.

To this end, suppose that there is a shock ε_t to the growth rate each period, and that this shock is generated by a stationary process with mean zero. Then a log-linear approximation to a discrete–time version of the generalized productivity-growth function (6) yields:

$$\Delta \ln A_t = \gamma_0 + \gamma_1 \left[\ln X_t - \ln Q_t + \left(\frac{\phi - 1}{\sigma} \right) \ln A_t \right] + \varepsilon_t \tag{10}$$

equation (10) applies to both theories, but with different parameter specification. If $\Delta \ln A_t$ were stationary then the expression in square brackets:

$$E_t = \ln X_t - \ln Q_t + \left(\frac{\phi - 1}{\sigma} \right) \ln A_t \tag{11}$$

would also have to be generated by a stationary process. Moreover, the results reported in column 1 of Table 4 below indicate that $\Delta \ln A_t$ is indeed stationary, confirming the impression of the previous section. That is, an augmented Dickey-Fuller test of the null hypothesis that $\Delta \ln A_t$ has a unit root is rejected with a p-value of 0.000. Hence if either of the two theories in question is correct then the variable representing expression E_t according to that theory must be stationary. This fact forms the basis of the cointegration tests reported below.

Table 4. Unit-root tests (Augmented Dickey-Fuller)

	Variable (X)	A	N (G5)	N (US)	R/A
$\ln X$	t-statistic	-1.03	-2.62	-1.90	-2.76
	p-value	0.736	0.095	0.328	0.072
$\Delta \ln X$	t-statistic	-7.00	-2.94	-2.87	-4.26
	p-value	0.000	0.048	0.057	0.001

A is US Productivity, N is the number of scientists and engineers engaged in R&D, and R is US real R&D expenditure. The sample period is 1953-2000 for R, and 1950-2000 for all others. See Appendix A for description of data. The null hypothesis is that $\ln X$ (resp. $\Delta \ln X$) contains a unit root. An intercept but no trend was included in the test equation. The lag length in the equation was chosen according to the Schwarz information criterion. Results were almost identical under the modified Schwarz criterion.

4.1 Semi-endogenous theory

For semi-endogenous theory, in which the product-proliferation effect is not present,[23] the factor Q_t is a constant, and the coefficient of $\ln A_t$ in (11) is strictly negative because the theory rests on the assumption that the degree of returns to scale ϕ is less than unity. So, given the stationarity of E_t, semi-endogenous theory implies that a linear combination of $\ln X_t$ and $\ln A_t$ must be stationary. More specifically, (a) the log of R&D input and log productivity must be integrated of the same order, and (b) if the two variables are nonstationary then they must be cointegrated, with a cointegrating vector $\left(1, \frac{\phi-1}{\sigma}\right)$ in which the second element is strictly negative.

The first column of Table 4 presents evidence to the effect that log productivity is I(1). That is, as already mentioned, the ADF test rejects the hypothesis that the first difference of log productivity has a unit root but it fails to reject the hypothesis that the level of log productivity has a unit root. Hence implication (a) above implies that the log of R&D input should be I(1). Columns 2 through 4 of Table 4 provide mixed evidence on this implication. Specifically, using all three measures of R&D input (G5 R&D workers, US R&D workers and productivity-adjusted R&D expenditure) the ADF test rejects a unit root in the first difference of log R&D input at the 10% significance level, but it also rejects a unit root in the log level of two of the three measures at the 10% level. However, in no case is a unit root in the log level rejected at the 5% level and in one case the p-value is close to 10%.

[23] We analyze the semi-endogenous framework with imperfect product proliferation in section 7 below.

If indeed implication (a) of semi-endogenous theory is true, then according to Table 4 both log productivity and log R&D input are most likely I(1) variables. Hence we can test implication (b) by carrying out a standard test of cointegration. Table 5 below reports the results of the Johansen test. The results indicate that we cannot reject the hypothesis that there is no cointegrating relationship between $\ln X$ and $\ln A$. Thus either implication (a) or implication (b) of semi-endogenous growth theory is not supported by the data.[24]

Table 5. Johansen cointegration tests for semi-endogenous theory

		trace statistic	max-eigenvalue statistic	cointegrating vector
$\ln N_{G5}$ and $\ln A$		11.73	8.19	$(1, 10.80)$ (5.09)
$\ln N_{US}$ and $\ln A$		10.01	7.29	$(1, -2.69)$ (0.41)
$\ln R/A$ and $\ln A$		12.82	9.58	$(1, -1.56)$ (0.51)
critical values	1%	20.04	18.63	
	5%	15.41	14.07	
	10%	13.33	12.07	

A is US productivity, N_{G5} and N_{US} are the number of scientists and engineers engaged in R&D in the G5 and the US respectively, and R is real US R&D expenditure. The sample period is 1950-2000 for N_{G5} and N_{US}, and 1953-2000 for R/A. See Appendix A for description of data. The null hypothesis is that the number of cointegrating equations is zero. An intercept but no trend was included in the cointegrating equation and the test VAR. Significance levels are denoted by * for 10%, ** for 5%, and *** for 1%. The small numbers in parentheses are standard errors. Critical values from Osterwald-Lenum (1992).

4.2 Schumpeterian theory

Schumpeterian theory is based on the assumption that the degree of returns to scale ϕ equals unity. Thus given the stationarity of the expression E_t defined by (11) above, the theory implies that the variable representing

$$F_t = \ln X_t - \ln Q_t$$

should be stationary. The expression F_t is the log of product-proliferation-adjusted R&D input, the variable represented by the measures displayed in Figures 7 and 8 above. Because F_t involves no free parameters, its stationarity can be tested directly using the Augmented Dickey-Fuller test. Table 6 below shows the results of such a test for each of our seven measures of adjusted R&D input. In all cases the hypothesis of a unit root in the log of adjusted R&D input is decisively rejected, thus corroborating the theory.

[24] It seems odd that the Johansen test reports that the best estimate of a cointegrating vector in the case where $X = N_{G5}$ attaches positive weight to both variables, given that both are trending upward. We attribute this to the fact that the test is looking for a non-existent vector.

Table 6. Unit-root tests (Augmented Dickey-Fuller) - adjusted R&D input

Variable (X)	N/L (G5)	N/L (US)	N/hL (G5)	N/hL (US)	R/AL	R/AhL	R/Y
t-statistic	-3.59	-3.06	-3.90	-3.46	-3.83	-3.62	-3.78
p-value	0.010	0.036	0.004	0.013	0.005	0.009	0.006

N is the number of scientists and engineers engaged in R&D, L is the number of employed US workers, h is human capital per US worker, R is US real R&D expenditure, A is US productivity and Y is US real GDP. The sample period is 1953-2000 for variables involving R, and 1950-2000 for all others. See Appendix A for description of data. The null hypothesis is that $\ln X$ contains a unit root. An intercept but no trend is included in the test equation. The lag length in the equation was chosen according to the Schwarz information criterion. The results were almost identical when the modified Schwarz criterion was used.

4.3 Summary

The cointegration analysis of this section does not constitute a decisive refutation of semi-endogenous theory, because failure to reject the hypothesis of no cointegrating relationship is not the same as rejection of the hypothesis that such a relationship exists. Moreover, as discussed in the previous section, it is possible that the period from 1950 to 2000 was a transitional period during which the long-term relationships predicted by semi-endogenous theory were masked by long lags that we are not taking into account.[25] However, while the data do not reveal the cointegrating relationships implied by stationary productivity growth under semi-endogenous theory for any of our empirical measures of gross R&D input, they do reveal the analogous cointegrating relationships implied under Schumpeterian theory for all of our empirical measures of adjusted R&D input. In this sense semi-endogenous theory does not seem to account for the long-run movements of R&D and productivity as well as fully-endogenous theory.

5 Forecasting the time series of productivity

As mentioned above, a key difference between semi-endogenous and Schumpeterian theory is the relative pessimism of the former's predictions concerning long-term growth. Accordingly one useful way of comparing the two competing theories is to see how well each does in forecasting productivity out of sample. As we shall see, semi-endogenous theory does much worse than Schumpeterian theory on this score, precisely because its forecasts are overly pessimistic; that is, semi-endogenous theory predicts that productivity growth should

[25] To see if this failure to detect the cointegrating relationships implied by semi-endogenous theory was attributable to transitional dynamics during just the first part of the sample period, when most of the decline in the growth rate of R&D input took place, we redid the cointegration tests three times, omitting the data before 1975, then before 1970 and then before 1965. In no case did this materially change the results reported in Table 5.

be falling in the final years of the 20th Century, whereas in fact, as we have seen, the trend during the last quarter of the 20th Century was for productivity growth to rise.

To predict we must first estimate the parameters of the two theories using data from a prior estimation period, and use the estimated parameter values to make out of sample forecasts. We perform the estimation as follows. A discrete-time version of the growth-equation (6) that applies to both theories is:

$$A_{t+1} = A_t + \lambda \left(X_t / Q_t \right)^{\sigma} A_t^{\phi}$$

In the case of semi-endogenous theory this amounts to:

$$A_{t+1} = A_t + \lambda X_t^{\sigma} A_t^{\phi}$$

and in second-generation fully endogenous theory it is:

$$A_{t+1} = A_t \left(1 + \lambda (X / Q_t)^{\sigma} \right)$$

In each case, we choose the parameters λ, σ and ϕ and an initial estimate $\widehat{A}_0$ to maximize the fit of a dynamic simulation to the observed series of log productivity,[26] under the restrictions imposed by the theory, namely that

$$0 \leq \sigma \leq 1 \text{ and } \phi < 1$$

for semi-endogenous theory and

$$0 \leq \sigma \leq 1 \text{ and } \phi = 1$$

for Schumpeterian theory. For semi-endogenous theory we used the number of scientists and engineers engaged in R&D in the G5 as the measure of R&D input, and for Schumpeterian theory we used the ratio of R&D expenditures to GDP as the measure of adjusted R&D input.

The first column of Table 7 below shows the result of this estimation when the entire sample period 1953-2000 is used as the estimation period. The first two rows indicate that semi-endogenous theory fits the time series a little better than its rival, with a standard error of 4.35 as compared with 4.49. However, because the fully-endogenous model contains the equality restriction $\phi = 1$, the semi-endogenous model has one more free parameter to fit the data. In order to make a comparison using the same number of degrees

[26] The procedure is equivalent to the nonlinear estimation procedure outlined by Jones (2002, Appendix A).

of freedom, we reestimated the semi-endogenous model adding a restriction on the ratio:

$$\gamma = \sigma / (1 - \phi),$$

which is the cointegration coefficient that should detrend the expression $\ln A_t - \gamma \ln N_t$ according to semi-endogenous theory. As Jones (2002) points out, an OLS regression of $\ln A_t$ on $\ln N_t$ should provide a superconsistent estimate of γ. This estimate is $\widehat{\gamma} = 0.263$. Accordingly, we fitted the semi-endogenous model with the additional parameter restriction: $\sigma = 0.263 \cdot (1 - \phi)$ in order to give it the same number of free parameters as the fully-endogenous model. As shown in the third row of Table 7 this resulted in a fit that was a little worse than that of the fully-endogenous model. As a point of reference, the fourth row shows the goodness of fit of neoclassical theory, represented by a linear trend:

$$A_{t+1} = A_t(1 + \eta)$$

where the only parameters to be estimated are η and $\widehat{A}_0$.

Table 7. Standard error of dynamic simulation of log productivity

	Basic simulation	Restricted simulation[a]	Restricted simulation[b]
Schumpeterian[c]	4.49	5.51	5.51
(ϕ, σ)	(1, 1)	(1, 1)	(1, 1)
Semi-endogenous[d]	4.35	5.40	5.63
(ϕ, σ)	(-4.9, 1)	(-5.2, 1)	(1, 0.07)
Restricted SE[e]	4.63	5.61	5.66
(ϕ, σ)	(-2.8, 1)	(1, 0)	(1, 0)
Neoclassical	4.67	5.61	5.66

(In parentheses are the estimated values of ϕ and σ.)
[a]Parameter estimates constrained to fit average productivity growth rate.
[b]Parameter estimates constrained to go through the data in 1953 and 2000.
[c]Using the R&D to GDP ratio as adjusted R&D input: X_t/Q_t
[d]Using G5 R&D scientists and engineers engaged in R&D as R&D input: X_t
[e]Subject to the restriction: $\sigma = 0.263 \cdot (1 - \phi)$, where 0.263 is the (superconsistent) OLS estimate of β in: $\ln A_t = \alpha + \beta \cdot \ln N_t$

By allowing the initial value of $\widehat{A}_0$ to be estimated we are allowing for the possibility that productivity is measured with error. However we are also allowing the fitted models to generate an average growth rate that could be quite different from what has been experienced. In fact, the basic simulations of all the above models predicted an overall change in productivity from 1953 to 2000 that was about 10 percent less than the actual change. Since growth theory is supposed to predict average long-run growth more so than the cyclical variation of productivity, we thought it appropriate to redo the data-fitting exercise this time constraining the parameter estimates of each model to result in an overall change in productivity equal to the observed overall change. The results are shown in the second column of Table 7. Similarly, the third column shows the result of fitting the data under the tighter constraint that the simulation goes through the actual productivity data in both 1953 and 2000. In all cases, it seems that neither theory fits the time series much better than the linear time trend of neoclassical theory.[27] Similar results were obtained when we pre-smoothed the productivity data using an H-P filter.

Table 8 shows the results of comparing the two theories' forecasting abilities. Specifically, it shows the root mean square forecast error of log productivity when the estimation by dynamic simulation, as described above, is conducted over an earlier sub-period and the results are used to forecast the remaining observations, conditional on the actual future values of the driving variables X_t and Q_t. (We allowed the value of $\sigma/(1-\phi)$ to be unconstrained by the OLS estimate.) As indicated in Table 8, the forecast error is substantially larger for semi-endogenous theory than for second generation fully endogenous theory in all cases but one. The average forecast error of semi-endogenous theory over the twelve cases reported was almost twice as large as the average for its Schumpeterian rival (8.51 versus 4.35).

Table 8. Root mean square forecast error in simulation of log productivity

	Schumpeterian[a]			Semi endogenous[b]		
Estimation period	Basic simulation	Restricted simulation[c]	Restricted simulation[d]	Basic simulation	Restricted simulation[c]	Restricted simulation[d]
1953-80	9.19	4.13	4.37	5.54	7.88	10.90
1953-85	3.07	3.22	3.60	7.98	8.94	7.98
1953-90	3.42	3.32	3.82	8.82	8.91	6.42
1953-95	5.43	3.15	5.46	11.20	10.68	6.83

[a]Using the R&D to GDP ratio as adjusted R&D input: X_t/Q_t
[b]Using G5 R&D scientists and engineers as R&D input: X_t
[c]Parameter estimates constrained to match the actual average productivity growth rate.
[d]Parameter estimates constrained to go through the data in 1953 and the last year of the estimation period.

[27]In three of the six simulations of semi-endogenous theory reported in Table 7, the best fit is obtained in the limiting case where the crucial returns to scale parameter ϕ equals unity, the one case not allowed by semi-endogenous theory. This is probably because the theory is trying to fit an inverse-U shaped curve to a U-shaped time series. As discussed above in section 2.2, the closest semi-endogenous theory can come to fitting a U-shaped time series occurs in the limiting case where $\phi = 1$, the case where the lag involved in transitional dynamics becomes infinitely long and the time path of productivity growth becomes a horizontal straight line.

More specifically, as indicated by our theoretical discussion of section 2.2.1 above, the semi-endogenous model in all cases predicts that productivity growth will continue to fall in the post sample period, whereas in fact the last quarter of the 20th Century was characterized by an upward trend of productivity growth. Figures 10 through 13 show the predictions of each theory, both in and out of sample, in four of the twelve cases – under the basic simulation and the (first) restricted simulation, when the estimation period ends respectively in 1980 and 1990. These figures are representative of all twelve cases, and they show that semi-endogenous theory is forecasting a productivity slowdown when productivity growth is in fact rising in the post-estimation period. Figure 10 illustrates the one case in which semi-endogenous theory has a smaller forecast error than Schumpeterian theory (the basic simulation when the estimation period ends in 1980). It shows that even in this case, by 2000 Schumpeterian theory is almost back on track while the semi-endogenous theory has again been increasingly underpredicting productivity for several years.

6 Using Civilian R&D expenditures

It might be argued that our measures of R&D are distorted by the large and rapid buildup of space and defense R&D during the 1950s and early 1960s. It is common in the IO productivity literature to remove these components on the grounds that their direct effects on productivity in the space and defense sectors cannot be measured for the absence of market prices, and because their spillovers into other sectors are hard to detect.

We did not do this in the tests reported above, on the grounds that to remove these components would be to remove over half of all R&D in some years. We believe that measures of R&D are too narrow to begin with, and that if anything we need much broader measures, to include innovative activities by small enterprises that are not included in the R&D statistics and to include activities even by large enterprises that are not done in a formal R&D lab but which contribute just as much to the creation of new technological knowledge.

Moreover, the failure of IO economists to detect cross-industry spillovers resulting from space and defense R&D does not demonstrate that such spillovers do not exist. On the contrary, the fact that so much of the modern computer and aeronautics industries have their origins in R&D sponsored by the US Department of Defense is direct evidence of the importance of non-civilian R&D for growth of aggregate productivity, especially in view of the fact that many analysts have attributed to the computer the lion's share of aggregate productivity growth through the last two decades of our sample period.[28]

[28] One reason why previous studies have failed to detect cross-industry spillovers from space and defense R&D may be that

Figure 10: Forecasts of log productivity - basic simulation - estimation period 1953-80

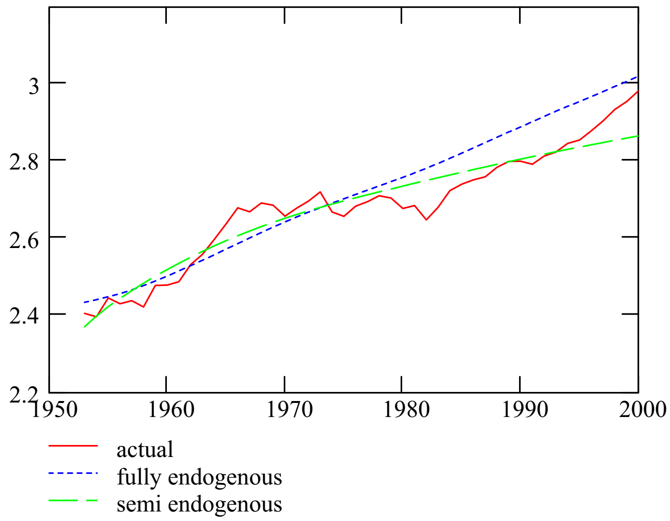


Figure 11: Forecasts of log productivity - basic simulation - estimation period 1953-90

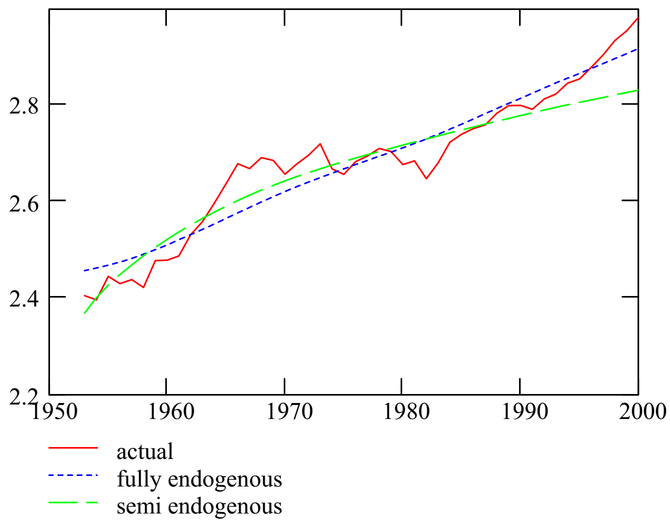


Figure 12: Forecasts of log productivity - restricted simulation - estimation 1953-80

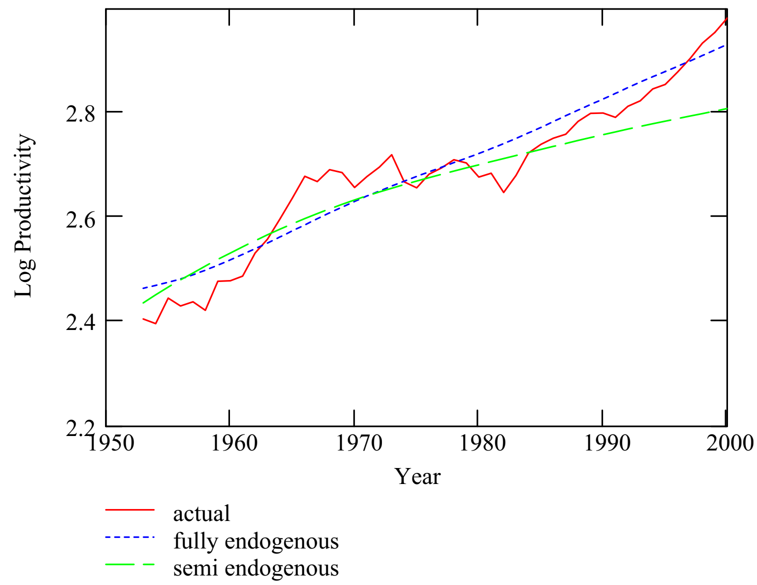


Figure 13: Forecasts of log productivity - restricted simulation - estimation 1953-90

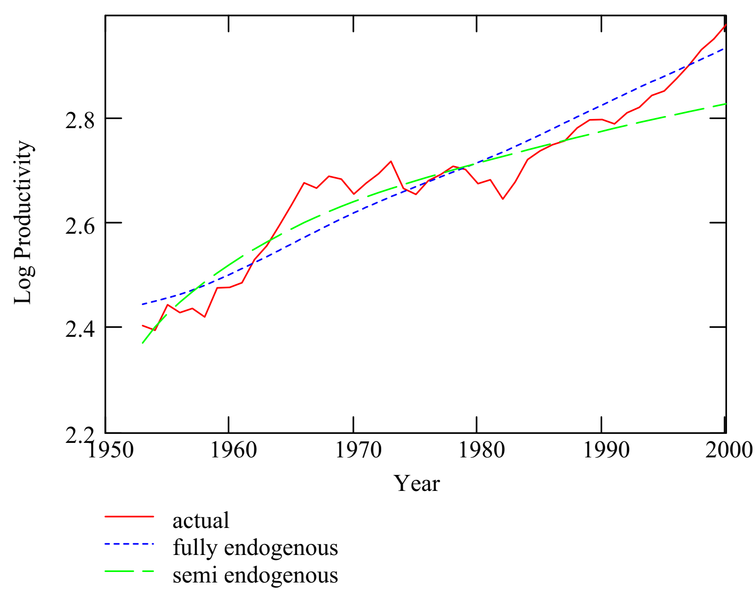
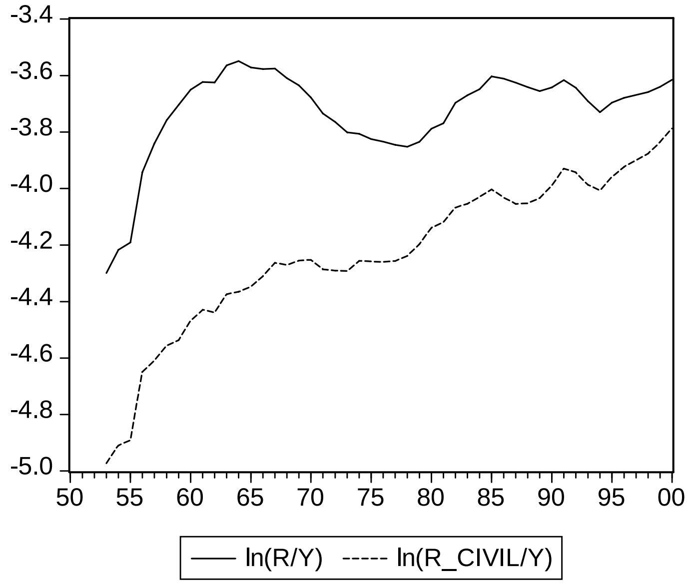


Figure 14: R&D to GDP ratio, total and civilian (log)



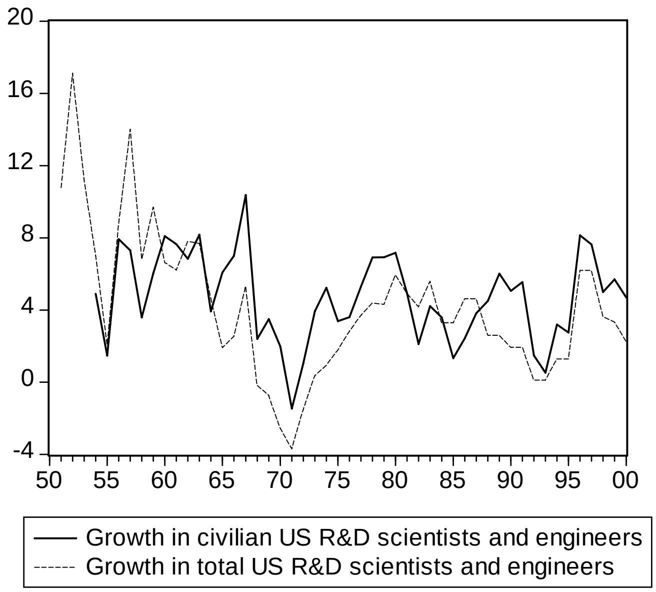
It is interesting nevertheless to consider what would happen if we were to remove the space and defense components from our measures of R&D. That is the purpose of the present section. Although this exercise does appear to remove some of the decline in growth of R&D input that is difficult for semi-endogenous theory to explain, and also creates an upward trend in the adjusted level of R&D inputs that would be difficult for Schumpeterian theory to account for, nevertheless it does not change the results of the cointegration tests reported above. Nor does the removal of space and defense resolve the problem of over-pessimism that semi-endogenous theory was confronted with in section 5 above; the theory still forecasts much worse than its Schumpeterian rival because it always predicts a post-sample productivity slowdown that has yet to materialize.

Figure 14 shows what we call "civilian" R&D expenditures (R_{civil})– total real US R&D expenditures R minus federal expenditures on space-related and defense-related R&D – expressed as a ratio to GDP, along with the ratio of total R&D to GDP The figure shows that, as indicated above, unlike the measure using total R&D, there is a steady upward trend in this adjusted R&D measure throughout second half of the 20th Century.

We created a rough estimate of the number of scientists and engineers engaged in civilian R&D (N_{civil}) by

they typically specify a detailed structure of cross-industry spillovers, which leaves too few degree of freedom to allow for long lags. Thus the spillovers from space and defense R&D may be too long and variable (across sectors as well as time) to show up in these studies. The results of Zachariadis (2003) indicate that when cross-industry spillovers are constrained to operate symmetrically across industries much longer lags are found than is typical in these studies. It would be interesting to revisit these studies to see if Zachariadis's strategy would reveal spillovers from space and defense R&D.

Figure 15: Growth rate in US R&D workers, total and civilian



multiplying the total number N by the ratio R_{civil}/R. Figure 15 shows that the growth rate of this measure of gross R&D input, instead of falling substantially over the sample period like the growth rate of total R&D workers, has a much smaller downward trend, which is fairly small in relation to both the variability and the average value of the series.

Table 9 shows that there is indeed no significant downward trend in the growth rate of civilian R&D workers, which tended to fall at almost the same rate as productivity growth, although the growth rate of our other gross measure of civilian R&D input, namely the productivity adjusted level of civilian R&D expenditures R_{civil}/A does still have a significant and substantial downward trend. Thus there is at least some support in the civilian R&D measures for semi-endogenous theory. Moreover, the last three columns of Table 9 show that all measures of product-proliferation-adjusted civilian R&D input have a significant and substantial upward trend during the sample period, in contradiction to Schumpeterian theory. In this respect, if civilian R&D were the right measure then semi-endogenous growth theory would actually be more consistent with observed trends than its Schumpeterian rival would be.

Table 9. Trends in productivity growth and various measures of civilian R&D

Variable (X)	g_A	$g_{N_{civil}}$	$g_{R_{civil}/A}$	R_{civil}/AL	R_{civil}/AhL	R_{civil}/Y
Average value[a]	1.40	4.73	4.78	4.99	2.19	1.51
Trend annual change[b]	-0.01	-0.03	-0.11	0.10	0.03	0.03
p-value[c]	0.614	0.301	0.087	0.000	0.000	0.000
Trend annual change[a] as percent of average value	-0.75	-0.68	-2.21	2.09	1.56	1.74

A is US Productivity, R_{civil} is US real non-space, non-defense R&D expenditure, N_{civil} is the number of scientists and engineers engaged in R&D multiplied by R_{civil} and divided by total US R&D expenditure, L is the number of employed US workers, h is human capital per US worker and Y is US real GDP. The sample period is 1950-2000 for g_A, and 1953-2000 for all others. See Appendix A for description of data.
[a]Expressed in percentage points
[b]OLS estimate of β in the regression: $X = \alpha + \beta \cdot year$
[c]Using HAC standard errors

However, when we revisit the cointegration tests of section 4 above using civilian R&D measures we find that they still favor Schumpeterian theory over semi-endogenous theory. Specifically, as Table 10 reports, there is still no evidence of the cointegrating relationship between productivity and gross R&D input that semi-endogenous theory predicts. Also, as Table 11 reports, the measures of adjusted R&D input that use R&D expenditures rather than R&D labor still pass the ADF test of stationarity as predicted by Schumpeterian theory.[29]

Table 10. Johansen cointegration tests for semi-endogenous theory

		trace statistic	max-eigenvalue statistic	cointegrating vector
$\ln N_{civil}$ and $\ln A$		4.46	3.19	$(1, -\underset{(0.69)}{4.36})$
$\ln R_{civil}/A$ and $\ln A$		9.12	6.79	$(1, \underset{(1.62)}{0.25})$
critical values	1%	20.04	18.63	
	5%	15.41	14.07	
	10%	13.33	12.07	

A is US productivity, R_{civil} is US real civilian R&D expenditure and N_{civil} is the number of scientists and engineers engaged in civilian R&D. The sample period is 1953-2000. See Appendix A for description of data. The null hypothesis is that the number of cointegrating equations is zero. An intercept but no trend was included in the cointegrating equation and the test VAR. Significance levels are denoted by * for 10%, ** for 5%, and *** for 1%. The small numbers in parentheses denote standard errors. Critical values from Osterwald-Lenum (1992).

[29]The rejection of non-stationarity in R_{civil}/Y is consistent with the upward trend apparent in Figure 14, because there is also an apparent slowdown in Figure 14, which seems to indicate that the ratio is slowly converging, like a sluggish but nonetheless stationary autoregressive process.

Table 11. Unit-root tests (Augmented Dickey-Fuller) - adjusted civilian R&D input

Variable (X)	N_{civil}/L	N_{civil}/hL	R_{civil}/AL	R_{civil}/AhL	R_{civil}/Y
t-statistic	-0.90	-0.76	-3.15	-3.29	-3.14
p-value	0.781	0.820	0.030	0.021	0.030

A is US Productivity, R_{civil} is US real civilian R&D expenditure, N_{civil} is the number of scientists and engineers engaged in civilian R&D, L is the number of employed US workers, h is human capital per US worker, and Y is US real GDP. The sample period is 1953-2000. See Appendix A for description of data. The null hypothesis is that ln$X$ contains a unit root. An intercept but no trend is included in the test equation. The lag length in the equation was chosen according to the Schwarz information criterion. Results were almost identical when the modified Schwarz criterion was used.

Table 12 shows that the forecasting ability of the semi-endogenous model, using the same procedure as in the previous section, is still much worse than that of Schumpeterian theory. The average forecast error of the semi-endogenous model in Table 12 is just over two times that of the Schumpeterian model. Examination of the detailed forecasts reveals that again the problem with the semi-endogenous forecasts is that they predict a productivity slowdown when the actual outcome has been a productivity speedup.

Table 12. Root mean square forecast error in simulation of log productivity using civilian R&D

	Schumpeterian[a]			Semi endogenous[b]		
Estimation period	Basic simulation	Restricted simulation[c]	Restricted simulation[d]	Basic simulation	Restricted simulation[c]	Restricted simulation[d]
1953-80	9.34	4.03	4.37	6.16	9.02	11.70
1953-85	3.04	3.02	3.60	9.15	10.24	8.61
1953-90	3.72	3.30	3.82	9.87	9.91	6.26
1953-95	5.90	3.76	5.46	10.80	10.43	6.68

[a] Using the civilian R&D to GDP ratio as adjusted R&D input: X_t/Q_t
[b] Using US scientists and engineers engaged in civilian R&D as R&D input: X_t
[c] Parameter estimates constrained to match the actual average productivity growth rate.
[d] Parameter estimates constrained to go through the data in 1953 and 2000.

To sum up the results of this section, although we believe there are strong *a priori* grounds for leaving the space and defense components in our measures of R&D input, it is not at all clear that omitting them would allow semi-endogenous theory to fit the trends in R&D and productivity better than Schumpeterian theory. For while it would give rise to an upward trend in adjusted R&D input that Schumpeterian theory would find hard to account for and would remove the downward trend in at least one of the measures of growth in R&D input that semi-endogenous theory would find hard to account for, nevertheless it does not resolve the other problems of semi-endogenous theory discovered above. Specifically, using civilian measures does not allow us to find the missing cointegration relationships of semi-endogenous theory or prevent us from finding those of Schumpeterian theory; nor does it stop the semi-endogenous model from making large forecast errors by predicting a falling growth rate throughout a twenty-year period of rising growth.

7 Double knife-edge and endogenous growth

The semi-endogenous model presented so far does not take into account the product proliferation effect. One can ask whether incorporating this effect would make any difference to the empirical validity of the model. In this section, we present an augmented version of the semi-endogenous model, and subject it to the same cointegration tests as in section 4 above. The results of these tests indicate that the data are unlikely to have been generated by a semi-endogenous model with product proliferation.

7.1 The semi-endogenous model with product proliferation

Although the semi-endogenous model presented in Section 2.2 specified $\beta = 0$ in equation (6), which characterizes the channel through which product proliferation affects the speed of technological progress, its distinctive characteristics would be preserved in the more general case where β is any number less than unity. In particular, Jones (2005) shows that when the semi-endogenous model is extended to this more general case, the number of different sectors increases with population but less than proportionally, and hence the model preserves the scale effect from population growth as the ultimate source of economic growth.

One example of such an augmented semi-endogenous model can be derived by assuming that $\beta < 1$ and $\phi < 1$ in equation (6), with $Q = L^{\beta}$ and $X = N$:

$$g_A = \lambda \left(\frac{N}{L^{\beta}} \right)^{\sigma} A^{\phi - 1} \tag{12}$$

Taking logs and differentiating both sides of equation (12) with respect to time, we have:

$$\frac{\dot{g}_A}{g_A} = \sigma \left(g_N - \beta n \right) - \left(1 - \phi \right) g_A$$

In the steady state, the number of R&D workers must be growing at the same rate as population growth

$(g_N = n)$ and g_A must be constant, so that we have:

$$\frac{\dot{g}_A}{g_A} = \sigma (1 - \beta) n - (1 - \phi) g_A = 0$$

from which the steady state value of g_A is:

$$g_A^* = \frac{\sigma (1 - \beta)}{1 - \phi} n$$

Thus in order to have long-run growth, we need to have either (a) $\phi < 1$ *and* $\beta < 1$ or (b) $\phi = 1$ *and* $\beta = 1$. Case (a) is an augmented semi-endogenous model, while case (b) is a limiting case that leads to the Schumpeterian model. The conditions for case (b) constitute what Jones (2005) calls a "double knife-edge," because two linearity conditions ($\phi = 1$ and $\beta = 1$) have to be satisfied simultaneously.

7.2 Cointegration analysis of the general model

Analogous to (10), a linear approximation to (12) in the case where R&D input is measured by the number of scientists and engineers in R&D (N) and Q equals the size of the labor force (L), yields:

$$\Delta \ln A_t = \gamma_0 + \gamma_1 \left[\ln N_t - \beta \ln L_t + \left(\frac{\phi - 1}{\sigma} \right) \ln A_t \right] + \varepsilon_t \qquad (13)$$

Assuming that all variables inside the square brackets are I(1), we can conduct cointegration tests using equation (13) and its modified versions with various definitions of X and Q. Considering the stationarity of $\Delta \ln A_t$, we need to have $\ln N$, $\ln L$, and $\ln A$ cointegrated with cointegrating vector $\left(1, -\beta, \frac{\phi-1}{\sigma}\right)$. If the semi-endogenous model is to be valid, β must be strictly less than unity and $\frac{\phi-1}{\sigma}$ must be strictly negative.

Table 13 shows the Johansen test results. According to these results, $\ln X$, $\ln L$, and $\ln A$ do indeed appear to be cointegrated. However, the estimated cointegrating vectors all imply that β is not less than one and $\frac{\phi-1}{\sigma}$ is not negative, both of which contradict the semi-endogenous approach. Although many of the coefficient estimates are imprecise, all but one lies one standard deviation or more away from the range implied by semi-endogenous theory, and a third of them are two standard deviations or more away.[30] Thus it seems that the data are not consistent with this more general version of semi-endogenous theory.

[30] Note also that the estimate of $\frac{\phi-1}{\sigma}$ could be any number if ϕ is close to one and σ is small, so results of Table 14 are consistent with the fully-endogenous specification: $\phi = 1$.

Table 13. Johansen cointegration tests for the augmented model

	trace statistic	max-eigen-value statistic	β	$\frac{\phi-1}{\sigma}$
$\ln N_{G5}, \ln L$, and $\ln A$	34.92**	30.34**	2.09 (0.34)	1.45 (0.70)
$\ln N_{G5}, \ln hL$, and $\ln A$	34.95**	28.77***	1.57 (0.29)	1.58 (0.76)
$\ln N_{G5}, \ln Y/A$, and $\ln A$	37.29***	28.90***	1.48 (0.27)	0.91 (0.64)
$\ln N_{US}, \ln L$, and $\ln A$	23.55	20.59*	2.25 (0.57)	1.85 (1.09)
$\ln N_{US}, \ln hL$, and $\ln A$	23.39	19.15*	1.73 (0.46)	2.01 (1.16)
$\ln N_{US}, \ln Y/A$, and $\ln A$	25.84	19.38*	1.66 (0.44)	1.43 (1.02)
$\ln R/A, \ln L$, and $\ln A$	45.36***	42.87***	1.60 (0.25)	0.97 (0.49)
$\ln R/A, \ln hL$, and $\ln A$	47.04***	43.51***	1.19 (0.19)	0.98 (0.48)
$\ln R/A, \ln Y/A$, and $\ln A$	48.73***	43.25***	1.10 (0.19)	0.55 (0.46)
critical values 1%	35.65	25.52		
critical values 5%	29.68	20.97		
critical values 10%	26.79	18.60		

Sample period is 1950-2000 for $\ln N_{G5}$ and $\ln N_{US}$, and 1953-2000 for $\ln R$. The null hypothesis is that the number of cointegrating equations is zero. An intercept but no trend was included in the cointegrating equation and the test VAR. It turned out that the number of cointegrating equations is at most one. Significance levels are denoted by * for 10%, ** for 5%, and *** for 1%. Small numbers in parentheses denote standard errors. Critical values from Osterwald-Lenum (1992).

7.3 Generality versus parsimony

Jones (2005) argues that the knife-edge linearity assumptions of $\phi = 1$ and $\beta = 1$ have little in the way of intuition or evidence to recommend them, and argues in favor of the more "general" model that can allow for a continuum of possible values of these parameters. However, by the same token, the double knife-edge assumption of Schumpeterian theory makes it more parsimonious than the more general semi-endogenous theory. Schumpeterian theory does not depend, for example, on long transitional dynamics to account for the data.[31] Being less tightly specified, semi-endogenous theory has more degrees of freedom with which to escape empirical refutation. The "double knife-edge" constitutes a pair of over-identifying restrictions which seem not to have worsened the empirical performance of Schumpeterian theory in comparison with the more

[31] But, there is a transitional dynamics leading to a long-run constancy in R&D intensity in the broader model that endogenizes capital accumulation and R&D. See Ha (2002).

general semi-endogenous theory. In this sense Schumpeterian theory appears to be both simpler than, and at least as powerful empirically as, semi-endogenous theory.

8 Conclusion

Two of the main objectives of growth theory are (a) to delineate the conditions needed in order to sustain long-term growth, and (b) to forecast long-term swings in growth rates. The analysis of this paper suggests that on both grounds, semi-endogenous growth theory does not perform as well as fully-endogenous theory, at least when it comes to accounting for the US experience over the second half of the 20th Century.

Concerning the first objective, over this period productivity growth has been trendless and stationary. According to the semi-endogenous model, sustaining a stationary trendless growth rate requires a trendless growth rate of R&D input and should result in a cointegrating relationship between productivity and the level of R&D input. Yet there has been a substantial downward trend in the growth rate of R&D input over the period, and the standard Johansen test fails to reveal such a cointegrating relationship. These results were found with all three of our measures of R&D input. In contrast, according to the second generation models of fully-endogenous growth theory, sustaining a stationary trendless growth rate requires a trendless and stationary share of resources allocated to R&D. Although there was a relatively small upward trend in the share according to some of the measures we used, two of them, both using R&D expenditures as the measure of gross R&D input, exhibited no significant upward trend, and all of our measures passed the standard ADF test of stationarity with a high degree of significance.

Concerning the second objective, the low frequency movements of TFP growth over the period exhibit something like a U-shaped pattern, with falling productivity growth over the period before 1975, especially since the early 1960s, and rising productivity growth since 1975. This corresponds roughly to the pattern exhibited by the share of resources allocated to R&D, especially when measured by the R&D to GDP ratio, as predicted by fully-endogenous theory. But it conflicts with the inverse-U shaped time path of productivity growth predicted by semi-endogenous theory given the falling growth rate of R&D input. Neither theory does much better than a linear time trend in explaining the time series of productivity, but the fact that semi-endogenous theory has been forecasting a downturn in productivity growth throughout the last two decades of the twentieth century, when in fact this has been a period of rising trend productivity growth, means that it does worse than fully endogenous theory, which has not predicted any such downturn.

These empirical failures of semi-endogenous theory relative to its fully-endogenous rival cannot be fully accounted for by removing the space and defense components from the R&D data, a practice which we would question in any event on *a priori* grounds. For although this tactic would remove some of the downward trend in R&D input that semi-endogenous theory finds hard to account for, and would also introduce a significant upward trend in the level of adjusted R&D input that Schumpeterian theory would find hard to account for, it does not change the relative failure of the semi-endogenous model with respect to cointegration tests or remove the problem of over-pessimism encountered by semi-endogenous theory in forecasting the time series on productivity.

None of these tests is decisive, especially given the relatively small number of observations for detecting long-run relationships and the possibility that the R&D production function whose form is at issue here might have shifted over the period.[32] Perhaps the difficulties of semi-endogenous theory could be explained by very long transitional dynamics in the relationship between R&D and productivity, or by a sequence of transitory shocks that have persisted over a half-century and masked the long-term relationships that we are trying to test. But, at the least, the fact that all of the tests seem to favor fully endogenous theory shows that the time series data on aggregate R&D support the theory's optimistic prediction that R&D has a permanent effects on growth as much as they support the more pessimistic semi-endogenous prediction of inevitable slowdown.

[32] Maddala and Kim (1999) discuss the problems that structural breaks pose to cointegration analysis.

Appendix A: Data

The data used in this paper are obtained and computed as follows.

TFP growth rates of the United States

TFP growth rates are computed through growth accounting exercises, which start from the following aggregate production function.

$$Y_t = K_t^{\alpha} \left(A_t h_t L_t\right)^{1-\alpha}$$

where Y_t is real GDP, K_t is capital stock, h_t is the level of human capital per worker, A_t is labor augmenting productivity index, and L_t is the number of workers. α has been assumed to be 0.3. Taking log and time derivatives, this function is transformed into the relationship between growth rates.

$$\frac{\dot{Y}_t}{Y_t} = \alpha \frac{\dot{K}_t}{K_t} + (1-\alpha)\left(\frac{\dot{A}_t}{A_t} + \frac{\dot{h}_t}{h_t} + \frac{\dot{L}_t}{L_t}\right)$$

This gives us the formula for TFP growth rates, $\dot{A}/A$.

$$\frac{\dot{A}_t}{A_t} = \frac{1}{1-\alpha}\frac{\dot{Y}_t}{Y_t} - \frac{\alpha}{1-\alpha}\frac{\dot{K}_t}{K_t} - \frac{\dot{h}_t}{h_t} - \frac{\dot{L}_t}{L_t}$$

We can compute $\frac{\dot{A}_t}{A_t}$ by plugging in the values of $\frac{\dot{Y}_t}{Y_t}$, $\frac{\dot{K}_t}{K_t}$, $\frac{\dot{L}_t}{L_t}$ and $\frac{\dot{h}_t}{h_t}$, each of which is obtained as follows.

First, $\frac{\dot{Y}_t}{Y_t}$ has been computed using the US real GDP data published by the Bureau of Economic Analysis, which is available online at http://www.bea.doc.gov/ bea/dn1.htm.

Second, the data for capital stock, K_t, have been obtained from the Bureau of Economic Analysis web site, http://www.bea.doc.gov/bea/dn/faweb/AllFATables.asp, Table 1.2. Chain-Type Quantity Indexes for Net Stock of Fixed Assets and Consumer Durable Goods, specifically the quantity indexes for nonresidential private and government fixed assets.

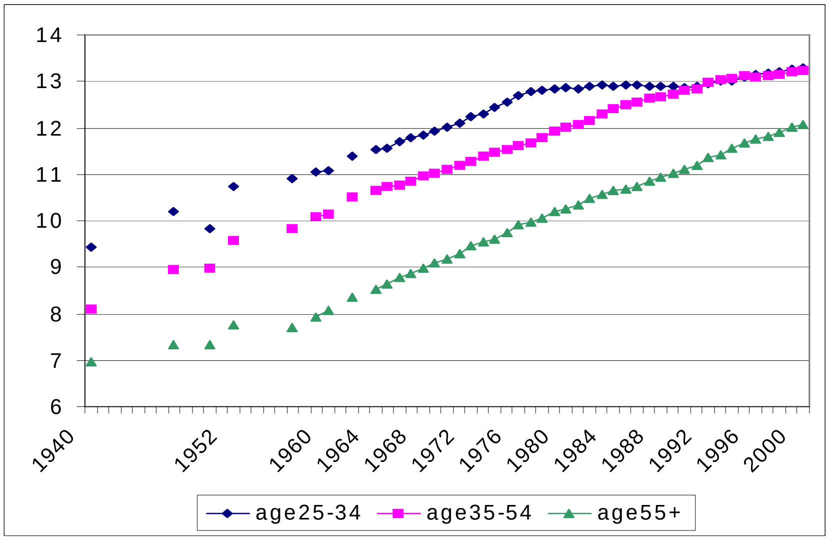
Third, the number of employees, L_t, has been obtained from the Bureau of Labor Statistics web site at ftp://ftp.bls.gov/pub/special.requests/ForeignLabor/flslforc.txt table 2.

Fourth, $\frac{\dot{h}_t}{h_t}$ has been computed by authors using the Mincerian approach, which is:

$$\begin{aligned} h_t &= e^{\theta s_t} \\ \frac{\dot{h}_t}{h_t} &= \theta \dot{s}_t \end{aligned}$$

We assumed that θ is 0.07, and computed s_t using the following method.

Figure 16: Average years of schooling for diffrerent age groups, (Source: Census Bureau and Authors' Calculation)



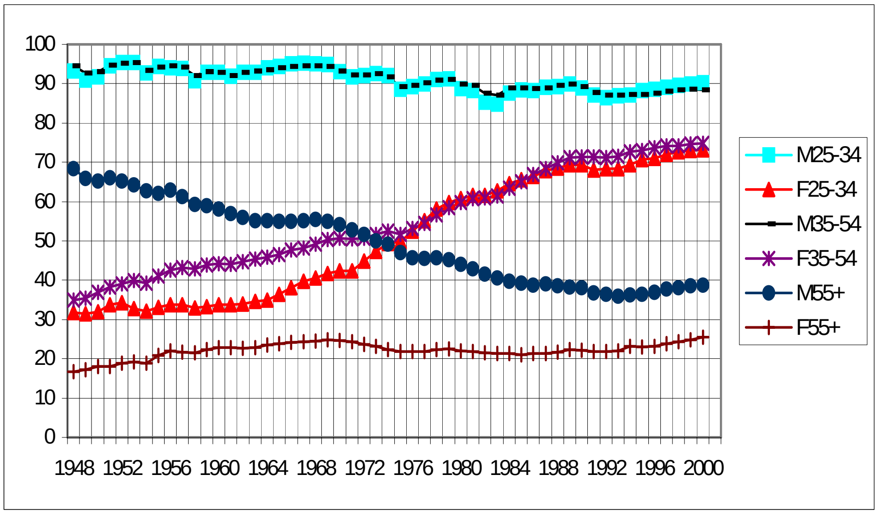
Years of schooling for US workers

Average educational attainment for US workers has been computed using various data from the Census Bureau and the Bureau of Labor Statistics. This task is more complicated than the cases for other data, since the average years of schooling for US workers, as opposed to the whole population, has not been compiled separately so far. We computed the average years of schooling for various age-sex subgroups and then obtained a weighted average for the employed workers.

First, the average years of schooling for each age-sex subgroups has been computed using the data from the Census Bureau, which is Table A-1 at http://www.census.gov/population/www/socdemo/educ-attn.html. This table reports the number of persons by cells of educational attainment. When computing the average, we assume that each person in a cell has the mean years of schooling for the cell, which is in turn computed using more detailed data in CPS tables linked on the same web site. Specifically, persons in the cell corresponding to zero to four years of schooling are assumed to have 1.8 years of schooling, 7.1 years is assigned for those who are in the cell 5 to 8 years of schooling, 10 years for 1 to 3 years of high school , 12 years for 4 years of high school, 13.8 years for 1 to 3 years of college, and 16 years for 4 years or more in college. Using this method, we computed the average years of schooling for the following six subgroups: Males age 25-34, Females age 25-34, Males age 35-54, Females age 35-54, Males age 55 and over, and Females age 55 and over. The average years of schooling for different age groups are depicted in Figure 16.

It is apparent that the average years of schooling for young people slowed down since the late 70s, while other age groups are reaching the top.

Figure 17: Employment-Population Ratio (%) of the US



Second, we computed each group's weight among the workers by using the data from the Bureau of Labor Statistics. Specifically, the number of employed people in each group is obtained from the data query process at http://data.bls.gov/cgi-bin/dsrv?lf, and the population of each group is obtained from the years of schooling table in the first step. Using those data, we computed the employment-population ratio for each group as in Figure 17. This tells us that the employment-population ratio of males age 55 and over has been dramatically decreasing probably due to the emergence of retirement and the increase of life expectancy. Also, young females, who have about the same level of education as males, experienced a significant increase in employment - population ratio.

Third, combining the above two steps, we computed the average years of schooling for workers according to the following formula.

$$s_t = \sum_i \frac{s_{it} L_{it}}{L_t}$$

where i represents each subgroup. s_t for missing years has been linearly interpolated, so that h_t is log-linearly interpolated. Now, we can see a slowdown of average years of schooling for employed people as shown in Figure 18, which implies that human capital per worker in the United States has not been growing constantly.

Considering the demographic changes, our method is closer to the reality than just simple average for the whole population. Figure 19 shows the differences between the average years of schooling for workers and that of whole population as well as the discrepancy from a linear trend.

Factor Accumulation Slowdown

It is worth mentioning that the trend of factor accumulation is far from steady. Combining the trends

Figure 18: Average Years of Schooling, US (Sources: Authors' calculation)

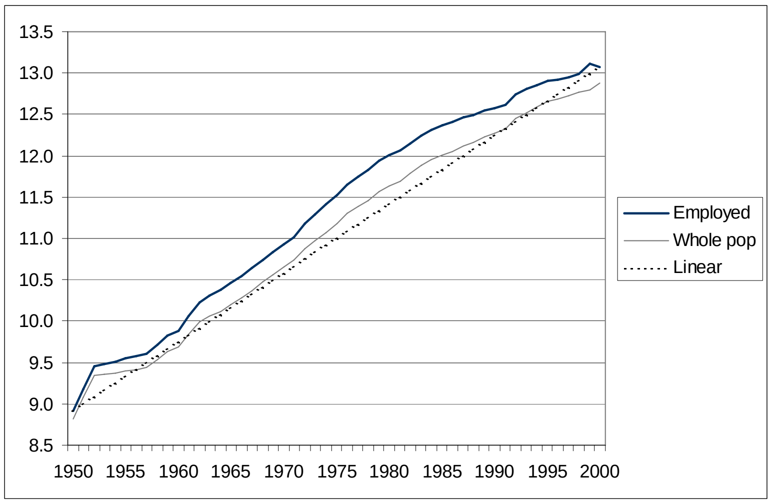


Figure 19: Decomposing the Difference between Average Years of Schooling for Employed and Linear Trend

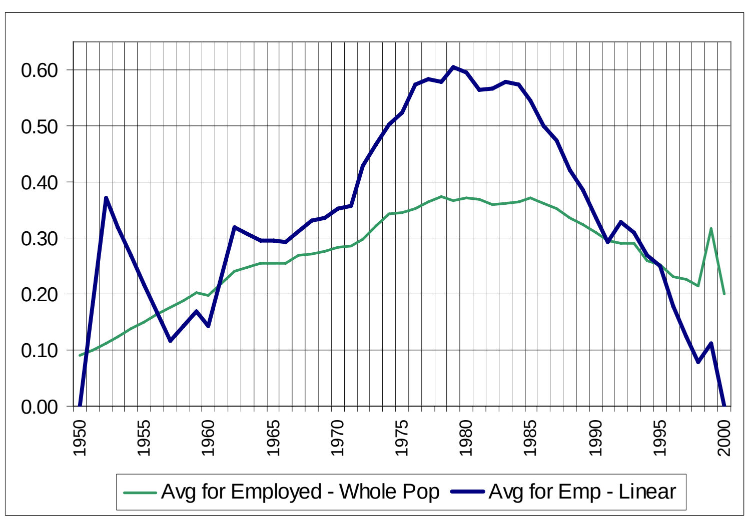
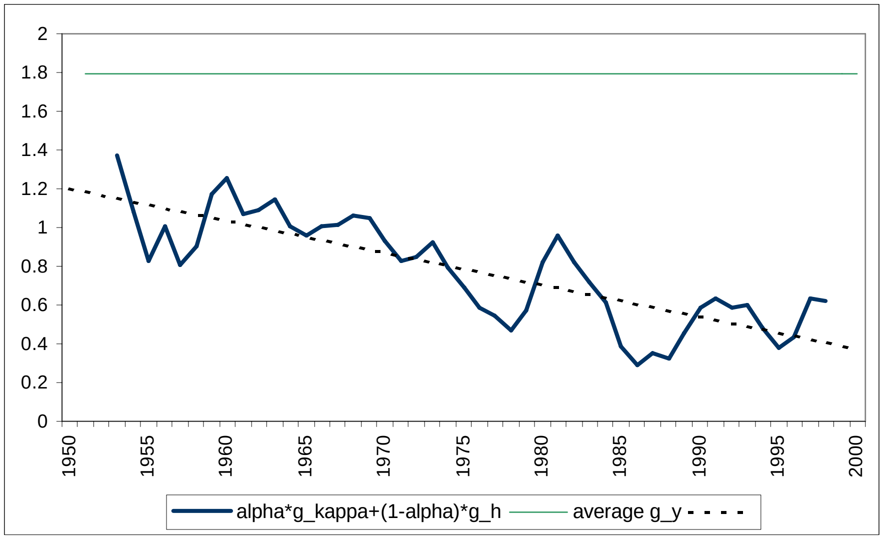


Figure 20: Factor Accumulation of the US, (5 year MA)



in human and physical capital accumulation, we can construct the factor accumulation trend of the US. Figure 20 shows that the rate of factor accumulation, $0.3\left(\frac{\dot{K}}{K}-\frac{\dot{L}}{L}\right)+0.7\frac{\dot{h}}{h}$ has a clear tendency of long-run slowdown. If a constant growth of output per worker is assumed, this implies that there might have been an acceleration of productivity growth in the long-run.

Scientists and Engineers Engaged in R&D

The data for 1950 to 1997 is from Science and Engineering Indicators 2000 published by the National Science Foundation, which can be obtained at http://www.nsf.gov/sbe/srs/seind00 in Appendix Table 3-25 and Charles Jones's web site (http://emlab.berkeley.edu/users/chad/Sources50.asc). Estimation for pre-1965 data for France, Germany, Japan, and UK has been made exactly the same way as in Jones (2002), where he assumes that the ratio of the number of R&D workers between these countries and the US in 1950 is the same as in 1965 and missing data are log-linearly interpolated. The number of G5 R&D workers for 1998-2000 has been estimated using the growth rate computed from the estimated number of total researchers for OECD in Table 07 of OECD Main Science and Technology Indicators 2002 published by OECD.[33][34] The number of US R&D workers for 1998-2000 has been estimated by assuming that the ratio of the number of R&D workers between G5 and the US in 1998-2000 is the same as in 1997.[35] Although there was a downward trend in the weight of the US among the G5 R&D workers, US's weight has been remarkably stable around

[33] OECD does not provide the estimates for every single country, but it does publish the estimates for OECD total. As G5 R&D workers constitute most of OECD R&D workers (76% in 1997), the growth rate of G5 R&D workers would be close to that of OECD R&D workers.

[34] The results of this paper are not sensitive to the inclusion of 1998-2000 data.

[35] Science and Engineering Indicators 2002 and 2004 do not provide the data for "scientists and engineers engaged in R&D" that can be directly comparable with the 1950-1997 data.

50% ever since the late 1970s.

R&D expenditure

The data for R&D as a percentage of GDP is obtained from the National Science Foundation web site at http://www.nsf.gov/sbe/srs/infbrief/nsf03307/figure2.xls. Total R&D expenditures are computed using this data and GDP data from BEA. Data on space-related and defense-related R&D as a percentage of total R&D are obtained from the same website at http://www.nsf.gov/sbe/srs/seind04/append/c4/at04-27.xls.

Appendix B: Accounting for the trend in the number of R&D workers

The trend in the number of R&D workers, N_t, can be written as follows:

$$N_t = \frac{N_t}{H_t}\frac{H_t}{L_t}L_t \tag{14}$$

where H_t is the number of skilled or highly educated workers, which can be defined in various ways. Taking log of equation (14), we have:

$$\ln N_t = \ln\frac{N_t}{H_t} + \ln\frac{H_t}{L_t} + \ln L_t$$

As $\ln L_t$ is roughly linear, we can guess that the logistic trend of $\ln N_t$ comes from either $\ln\frac{N_t}{H_t}$ or $\ln\frac{H_t}{L_t}$, or both. Figure 21 shows that various definitions of $\frac{N_t}{H_t}$ have no monotonic trend but an inverse-U shaped trend with a hump around 1964. It is notable that R&D workers as a percentage of workers with 12, 13, and 16 years of schooling all show a slight downward trend in the post-1964 period.[36]

On the other hand, in Figure 22 the log trend of $\frac{H_t}{L_t}$ seems to be close to linear, but with a careful look, one can detect a slowdown of the slopes in the later phase.

Therefore, the logistic growth of the number of R&D workers can be explained by the downward trend of R&D workers to educated workers ratios in the last 35 years of the 20th Century together with the logistic growth of educational attainment.[37]

[36] As the time series for H is too short (from 1992), H has been estimated by adding the number of educated workers for various age-sex cohorts with the employment-population ratio being the weight for each group.

[37] Figure 18 above.

Figure 21: R&D workers to educated workers ratios (log, age 25 and over), US

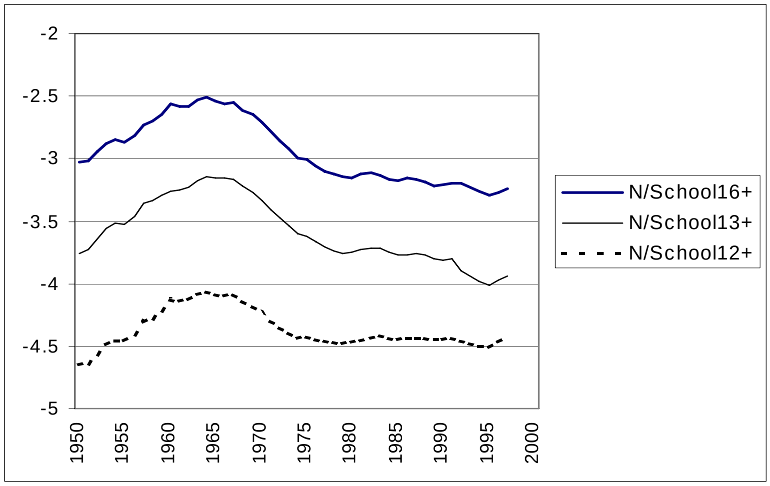
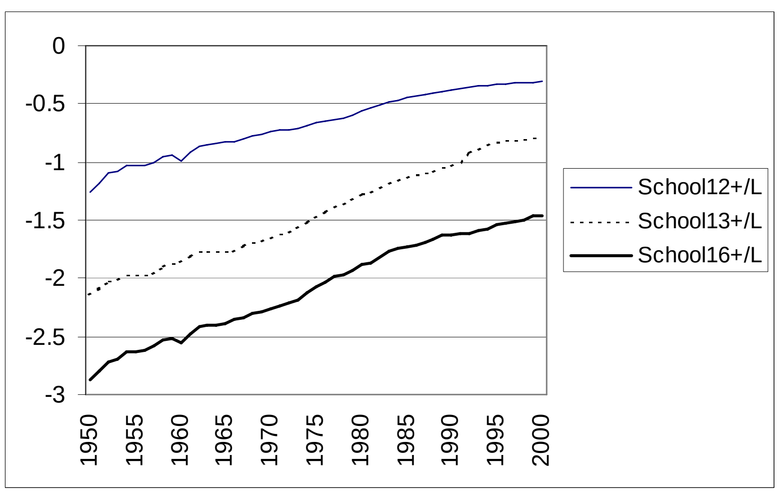


Figure 22: Trend of the proportion of educated workers (log, age 25 and over), US



References

Aghion, P., Bloom, N., Blundell, R., Griffith, R. and Howitt, P. (2005), "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics* 120: 701-28.

Aghion, P. and Howitt, P. (1992), "A Model of Growth Through Creative Destruction," *Econometrica* 60: 323-351.

Aghion, P. and Howitt, P. (1998a), *Endogenous Growth Theory*, MIT Press.

Aghion, P. and Howitt, P. (1998b), "On the Macroeconomic Effects of Major Technological Change," in Helpman, E. ed. *General Purpose Technologies and Economic Growth*, MIT Press.

Aghion, P., Howitt, P. and Mayer-Foulkes, D. (2005), "The Effect of Financial Development on Convergence: Theory and Evidence," *Quarterly Journal of Economics* 120: 173-222.

Arora, S. (2001), "Health, Productivity, and Long-Term Economic Growth," *Journal of Economic History* 61: 699-749.

Baily, M. N. and Lawrence, R. Z. (2001), "Do We Have a New E-conomy?" *American Economic Review Papers and Proceedings* 91: 308-12.

Barro, R. J. and Sala-i-Martin, X. (1995), *Economic Growth*, McGraw-Hill.

Basu, S., Fernald, J. G., and Shapiro, M. D. (2001), "Productivity Growth in the 1990s: Technology, Utilization, or Adjustment?" *Carnegie-Rochester Conference Series on Public Policy* 55: 117-65.

Coe, D. T. and Helpman, E. (1995), "International R&D Spillovers," *European Economic Review* 39: 859-887.

Corriveau, L. (1991), "Entrepreneurs, Growth, and Cycles," Ph.D. Dissertation, University of Western Ontario.

Dinopoulos, E. and Thompson, P. (1998), "Schumpeterian Growth without Scale Effects," *Journal of Economic Growth* 3: 313-335.

Easterly, W. and Levine, R. (2001), "It's Not Factor Accumulation: Stylized Facts and Growth Models," *World Bank Economic Review* 15: 177-219.

Gordon, R. (2002), "Technology and Economic Performance in the American Economy," *NBER Working Paper* 8771.

Griliches, Z. (1994), "Productivity, R&D, and the Data Constraint," *American Economic Review* 84: 1-23.

Grossman, G. M. and Helpman, E. (1991), "Quality Ladders in the Theory of Growth," *Review of Economic Studies* 58: 43-61.

Ha, J. (2002), "From Factor Accumulation to Innovation: Sustained Economic Growth with Changing Components," mimeo.

Howitt, P. (1999), "Steady Endogenous Growth with Population and R&D Inputs Growing," *Journal of Political Economy* 107: 715-730.

Howitt, P. (2000), "Endogenous Growth and Cross-Country Income Differences," *American Economic Review* 90: 829-846.

Howitt, P. and Aghion, P. (1998), "Capital Accumulation and Innovation as Complementary Factors in Long-Run Growth," *Journal of Economic Growth* 3: 111-130.

Howitt, P. and Mayer-Foulkes, D. (2005), "R&D, Implementation and Stagnation: A Schumpeterian Theory of Convergence Clubs," *Journal of Money, Credit, and Banking* 37: 147-77.

Jones, C. I. (1995a), "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics* 110: 495-525.

Jones, C. I. (1995b), "R&D-Based Models of Economic Growth," *Journal of Political Economy* 103: 759-784.

Jones, C. I. (2002), "Sources of U.S. Economic Growth in a World of Ideas," *American Economic Review* 92: 220-239.

Jones, C. I. (2005), "Growth and Ideas," in Aghion. P. and Durlauf, S. N. eds. *Handbook of Economic Growth,* North-Holland.

Kortum, S. (1997), "Research, Patenting, and Technological Change," *Econometrica* 65: 1389-1419.

Laincz, C. A., and Peretto, P. F. (2004), "Scale Effects in Endogenous Growth Theory: An Error of Aggregation Not Specification," mimeo.

Maddala, G. S., and In-Moo Kim. (1999) *Unit Roots, Cointegration, and Structural Change,*Cambridge University Press.

Maddison, A. (2001), *The World Economy: A Millennial Perspective*, OECD.

OECD, *Main Science and Technology Indicators 2002*, OECD.

Osterwald-Lenum, M. (1992), "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics* 54: 461-72.

Pack, H. (1994), "Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings," *Journal of Economic Perspectives* 8: 55-72.

Parente, S. L. and Prescott, E. C. (1999), "Monopoly Rights: A Barrier to Riches," *American Economic Review* 89: 1216-1233.

Peretto, P. (1998), "Technological Change and Population Growth," *Journal of Economic Growth* 3: 283-311.

Rivera-Batiz, L. A. and Romer, P. M. (1991), "Economic Integration and Endogenous Growth," *Quarterly Journal of Economics* 106: 531-555.

Romer, P. M. (1990), "Endogenous Technological Change," *Journal of Political Economy* 98: 71-102.

Segerstrom, P. S. (1998), "Endogenous Growth Without Scale Effects," *American Economic Review* 88: 1290-1310.

Segerstrom, P. S., Anant, T. C. A., and Dinopoulos, E. (1990), "A Schumpeterian Model of the Product Life Cycle," *American Economic Review* 80: 1077-1091.

Ulku, H. (2005), "An Empirical Analysis of R&D-Based Growth Models," mimeo, Manchester University.

Young, A. (1998), "Growth without Scale Effects," *Journal of Political Economy* 106: 41-63.

Zachariadis, M. (2003), "R&D, Innovation, and Technological Progress: A Test of the Schumpeterian Framework without Scale Effects," *Canadian Journal of Economics* 36: 566-86.

Zachariadis, M. (2004), "R&D-Induced Growth in the OECD?" *Review of Development Economics* 8: 423-39.