Semi-endogenous versus Schumpeterian growth models: a critical review of the literature and new evidence

Herzer, Dierk

May 2020

Semi-endogenous versus Schumpeterian growth models: a critical review of the literature and new evidence

Dierk Herzer*

Department of Economics, Helmut-Schmidt-University Hamburg, Holstenhofweg 85, 22043 Hamburg, Germany

ABSTRACT

Several studies have tested semi-endogenous versus Schumpeterian growth models using different methodological approaches. This paper critically reviews these studies including their approaches and provides new evidence on this issue, by analyzing both time-series data from the United States and panel data from 19 OECD countries over the period 1980-2014. The review finds much support for Schumpeterian growth theory, but shows that all studies reviewed have several limitations, including conceptual problems associated with the use of the number/stock of patents as a measure of the flow/stock of knowledge, the possibility of spurious regressions due to non-stationary data, potential mismeasurement of R&D inputs due to possible interpolation and deflation errors, misspecification problems that can arise in difference models when variables are cointegrated, and potential spurious rejections of the unit root hypothesis for R&D intensity when the lag length in unit root tests is too small. The present study avoids these limitations and finds strong evidence in favor of semi-endogenous growth.

JEL classification:

O30, O40, O11

Keywords:

semi-endogenous growth models, Schumpeterian growth models, R&D, TFP, unit roots, cointegration

* *E-mail address:* herzer@hsu-hh.de

1. Introduction

Most recent research and development (R&D) based models of economic growth fall into two broad categories. On the one hand, there are Schumpeterian growth models (see, e.g., Young, 1998; Peretto, 1998; Howitt, 1999), in which new knowledge is produced using resources devoted to R&D and the existing stock of knowledge that exhibits constant returns. As an economy grows, however, the associated proliferation of product varieties reduces the effectiveness of R&D because a given amount of R&D resources has to be spread more thinly over a larger number of varieties. Since the number of varieties, in the long run, is proportional to the scale of the economy, the growth rate of the stock of knowledge, and thus the growth rate of output per capita, depends on the amount of resources devoted to R&D relative to the scale of the economy, commonly referred to as R&D intensity. On the other hand, there are semi-endogenous growth models (see, e.g., Jones, 1995; Kortum, 1997; Segerstrom, 1998), which assume diminishing returns to the stock of knowledge and absence of negative product proliferation effects. Semi-endogenous growth models therefore imply that the growth rates of knowledge and per capita output depend on the growth rate of research effort.[1] Finally, Jones (1999) considers a hybrid semi-endogenous model with diminishing returns to the stock of knowledge and partial product proliferation. In such a model, knowledge growth depends negatively on growth in the scale of the economy and positively on growth in research effort.

Several studies have empirically tested these theories against each other using different methodological approaches. The contributions of this paper lie in (1) critically reviewing these studies and their approaches and (2) providing new evidence on the validity of these theories.

Studies that have tested either Schumpeterian theory or semi-endogenous growth theory are also discussed but are not explicitly included in the review because they are by their nature unable to reject the other, untested theories, whose predictions overlap in part with those of the tested theories. As we demonstrate in this paper, confirmatory evidence in these studies for the tested theory therefore does not necessarily imply the validity of the theory.

Our review finds much support for Schumpeterian growth theory, but shows that all of these studies suffer from at least one of several limitations, including conceptual problems associated with the use of the number/stock of patents as a measure of the flow/stock of knowledge, the possibility of spurious regressions due to non-stationary data, possible measurement error

[1] Semi-endogenous growth models assume that, along a balanced growth path, the growth rate of research effort is equal to the growth rate of the population, implying that in a balanced growth equilibrium, the growth rates of knowledge and per capita output are proportional to the growth rate of the population. However, out of steady state, when the number of researchers and the size of the population do not grow at the same rate, the growth rates of knowledge and per capita output are only proportional to the growth rate of research effort, as discussed in Jones (2002) (and later in this paper). In this sense, and explicitly noted by Jones (1995, p. 777), "[t]he [semi-endogenous growth] model actually relates the growth in the effective number of researchers, rather than in population, to economic growth."

associated with the use of interpolated R&D input data and/or the use of an overall price deflator to deflate R&D expenditures, misspecification problems that arise when difference models with cointegrated variables are estimated without an error-correction term, and potential spurious rejections of the unit root hypothesis for R&D intensity when the lag length in unit root tests is too small.

To address these limitations, we present new empirical evidence on the validity of R&D-based models of growth, using total factor productivity (TFP) as a measure of the stock of knowledge, employment as a measure of scale, and non-interpolated data on the number of researchers as a measure of research effort within a unit-root/cointegration framework for non-stationary data combined with proper lag selection procedures.

To increase the comparability of our results with those in the existing literature, which focuses mainly on the United States and other relatively advanced economies, we conduct both a time-series analysis for the United States and a panel data analysis of 19 OECD countries; the United States is excluded from the panel analysis to ensure that the panel results are not driven by the observations of the United States. To preview our main result, we find empirical support for semi-endogenous growth theory in explaining recent performance in the OECD countries, including the United States.

The remainder of this paper is structured as follows. Section 2 provides the theoretical framework for our review of the empirical literature.[2] The literature is reviewed in Section 3. Section 4 presents the new empirical evidence, and Section 5 concludes.

2. Schumpeterian and semi-endogenous growth models

2.1. General specifications

We assume an aggregate Cobb-Douglas production function of the form

$$Y_t = A_t K_t^{\alpha} (L_t h_t)^{1-\alpha},\ 0 < \alpha < 1, \tag{2.1}$$

where Y_t is output at time t; K_t denotes the stock of physical capital; $L_t h_t$ is the amount of human capital-augmented labor used in production, defined as the number of workers, L_t, times human capital per worker, h_t; and A_t represents the stock of knowledge. The production function in (2.1) can be rewritten in terms of output per worker $y_t \equiv Y_t/L_t$ as

[2] Bond-Smith (2019) provides a comprehensive survey of the theoretical work that has been carried out on endogenous growth over the last thirty years. In Section 2, we focus on the theoretical issues relevant to our review of the empirical literature that has tested semi-endogenous against Schumpeterian growth models.

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

(2.2)

The central component of any R&D-based growth model is a knowledge production function. A general form of this function which allows us to distinguish between semi-endogenous and Schumpeterian growth theories is (see, e.g., Ha and Howitt, 2007; Ang and Madsen, 2011)

$$\dot{A}_t = \delta A_t^{\phi} \left(\frac{X_t}{Q_t^{\beta}}\right)^{\lambda}$$

(2.3)

where $\dot{A}_t$ is the flow of new ideas or knowledge produced at time t; δ is a constant of proportionality; ϕ is a parameter that describes the nature of returns to scale in the production of new ideas, X_t is R&D input, usually measured by R&D labor or R&D expenditure; λ, where $0 < \lambda \leq 1$, is a parameter that captures the extent to which an increase in R&D effort induces duplication; Q_t is the number of product varieties or sectors, which, as part of the denominator of the last term in (2.3), captures possible product proliferation effects that reduce the effectiveness of R&D; finally, β is the parameter of product proliferation.

Schumpeterian models of economic growth are based on the assumption that $\beta = 1$ and $\phi = 1$. In these models, the variable Q_t, which can be interpreted more generally as a scale variable, is proportional to the size of the labor force L_t (or the population) and hence can depend on any variable that grows in the long run at the same rate as the labor force, such as output. Semi-endogenous growth models are characterized by the absence of product proliferation, $\beta = 0$, and diminishing returns to knowledge, $\phi < 1$. Finally, Jones's (1999) hybrid semi-endogenous model considers the case where $0 < \beta < 1$ and $\phi < 1$.

In all these models, the long-run growth rate of output per capita/worker depends on the long-run growth rate of the knowledge stock. Rewriting (2.3) in terms of the growth of the stock of knowledge gives

$$\frac{\dot{A}_t}{A_t} = \delta A_t^{\phi-1} \left(\frac{X_t}{Q_t^{\beta}}\right)^{\lambda}$$

(2.4)

2.2. Schumpeterian growth models

If $\phi = 1$ and $\beta = 1$, then equation (2.4) becomes

$$\frac{\dot{A}_t}{A_t} = \delta\left(\frac{X_t}{Q_t}\right)^{\lambda} \tag{2.5}$$

This equation predicts that long-run changes in R&D intensity, X_t/Q_t, lead to long-run changes in knowledge growth. If the growth rate of knowledge is constant, then the above equation implies that R&D intensity must also be constant.

The logic behind equation (2.5) can be described as follows (see, e.g., Jones, 1999; Laincz and Peretto, 2006). Assuming that the production function (2.1) is an average production function for varieties (that are produced in the same quantity) and that aggregate output is given by

$$Y_t = Q_t^{\theta} A_t K_t^{\alpha} (L_t h_t)^{1-\alpha} \tag{2.6}$$

we can express per capita output as the product of $Q_t^{\theta-1}$ and the average sectoral output per sectoral worker $A_t^{\frac{1}{1-\alpha}}\left(\frac{K_t}{Y_t}\right)^{\frac{\alpha}{1-\alpha}} h_t$,

$$y_t = Q_t^{\theta-1} A_t^{\frac{1}{1-\alpha}}\left(\frac{K_t}{Y_t}\right)^{\frac{\alpha}{1-\alpha}} h_t \tag{2.7}$$

where $\theta > 1$ is related to the elasticity of substitution between products and captures the variety effect—that is, the increased productivity that results from the increased availability of specialized inputs. Along a balanced growth path, all terms on the right side of (2.7) except for $Q_t^{\theta-1}$ and $A_t^{\frac{1}{1-\alpha}}$ are constant. Thus, the growth rate of per capita/worker output in steady-state is given by

$$g_y = (\theta - 1) g_Q + \frac{1}{1-\alpha} g_A \tag{2.8}$$

where g_Q is the steady-state growth rate of varieties and g_A is the steady-state growth rate of *average* knowledge. With a sectoral knowledge production of the form

$$\dot{A}_t = \delta A_t X_t^{\lambda}$$

(2.9)

the steady-state growth rate of A therefore depends on research effort *per sector* and thus on R&D intensity,

$$g_A = \lambda\delta\frac{X}{Q}$$

(2.10)

This is the logic behind equation (2.5), which is based on the idea that an increase in scale increases the potential supply of R&D inputs and the number of products available in the same proportion, leaving R&D intensity unchanged.

Substituting (2.10) into (2.8) yields

$$g_y = (\theta - 1)g_Q + \frac{\lambda\delta}{1-\alpha}\frac{X}{Q}$$

(2.11)

Since in a balanced growth equilibrium the growth rate of the labor force g_L is equal to the growth rate of varieties, we can also write the steady-state growth rate of per capita/worker output as

$$g_y = (\theta - 1)g_L + \frac{\lambda\delta}{1-\alpha}\frac{X}{Q}$$

(2.12)

This equation predicts that the growth rate of output per capita/worker is positively related to the level of R&D intensity and the growth rate of the labor force (or the population). The latter requires that $\theta > 1$. If $\theta = 1$, then population or employment growth does not affect steady-state growth, and steady-state growth depends only on R&D intensity (see, e.g., Laincz and Peretto, 2006).

2.3. Semi-endogenous growth models

If $\phi < 1$ and $\beta = 0$, then equation (2.4) reduces to

$$\frac{\dot{A}_t}{A_t} = \delta A_t^{\phi-1} X_t^{\lambda}$$

(2.13)

Assuming that the stock of knowledge grows in the long run at a constant rate g_A, the above equation can be solved for the stock of knowledge, yielding

$$A_t = \left(\frac{\delta}{g_A}\right)^{\frac{1}{1-\phi}} X_t^{\frac{\lambda}{1-\phi}}$$

(2.14)

It follows from (2.14) that $X_t^{\frac{\lambda}{1-\phi}}$ is proportional to A_t if the growth rate of knowledge is constant. In this case, the long-run growth rate of the knowledge stock is given by

$$g_A = \frac{\lambda}{1-\phi} g_X$$

(2.15)

where g_X is the long-run growth rate of research effort.

Along a balanced growth path, g_X is equal to g_L (or g_Y), implying that X_t is proportional to L_t (or Y_t). In a balanced growth equilibrium, equation (2.14) therefore implies

$$A_t = \left(\frac{\delta}{g_A}\right)^{\frac{1}{1-\phi}} L_t(\text{or } Y_t)^{\frac{\lambda}{1-\phi}}$$

(2.16)

and

$$g_A = \frac{\lambda}{1-\phi} g_{L\,(\text{or } Y)}$$

(2.17)

Thus, as can be seen by rewriting equation (2.4) as

$$\frac{\dot{A}_t}{A_t} = \delta \left(\frac{X_t}{Q_t}\right)^{\lambda} \left(\frac{A_t}{Q_t^{\frac{\lambda}{1-\phi}}}\right)^{\phi-1}$$

(2.18)

both R&D intensity is constant and $Q_t^{\frac{\lambda}{1-\phi}}$ is proportional to A_t in a balanced growth equilibrium. Moreover, comparing equation (2.18) with (2.5), it is apparent that both semi-endogenous theory and Schumpeterian theory are consistent with a constant ratio of L_t (or Y_t) to Q_t and a constant growth rate of knowledge/output per worker, but semi-endogenous theory (Schumpeterian theory) requires that in steady state $Q_t^{\frac{\lambda}{1-\phi}}$ is (not) proportional to A_t.

It is important to emphasize here that equations (2.16) and (2.17) are valid only if the share of labor devoted to R&D (or the stare of output going to R&D) is constant over time. However, to the extent that the amount of resources allocated to R&D depends on additional factors beyond the mere size of the labor force (such as education, international competition, and access to foreign markets) and to the extent that these additional factors change over time, they induce changes over time in X_t, so that L_t (or Y_t) and X_t do not grow at the same rate over long periods of time.

Jones (2002) therefore distinguishes between a constant growth path and a balanced growth path. Along both paths, growth rates are constant, but the latter is associated with a steady state or balanced growth equilibrium in which all variables in non-per capita terms grow the same rate forever, so that R&D intensity is necessarily constant. The implication is that, in semi-endogenous growth theory, in contrast to Schumpeterian growth theory, R&D intensity may not be constant over long time intervals even though the growth rate of knowledge is constant. To see this, rewrite equation (2.14) as

$$A_t = \left(\frac{\delta}{g_A}\right)^{\frac{1}{1-\phi}} \left(\frac{X_t}{L_t\,(\text{or } Y_t)}\right)^{\frac{\lambda}{1-\phi}} L_t\,(\text{or } Y_t)^{\frac{\lambda}{1-\phi}} \tag{2.19}$$

Taking logs and differencing the equation yields

$$g_A = \frac{\lambda}{1-\phi} g_{(X/L \text{ or } Y)} + \frac{\lambda}{1-\phi} g_{L\,(\text{or } Y)} \tag{2.20}$$

where $g_{(X/L \text{ or } Y)}$ is the growth rate of R&D intensity. Thus, out of steady state, the growth rate of knowledge can be constant and depend on both the growth rate (rather than the level) of R&D intensity and the growth rate of the labor force (or the growth rate of output). In steady state, the first term in equation (2.20) must be zero, so that this equation reduces back to equation (2.17). In this case, it follows from the steady-state version of equation (2.2),

$$g_y = \frac{1}{1-\alpha} g_A \tag{2.21}$$

that the growth rate of output per capita/worker g_y is given by

$$g_y = \frac{\lambda}{(1-\alpha)(1-\phi)} g_L$$

(2.22)

or

$$g_y = \frac{\lambda}{(1-\alpha)(1-\phi)} g_X$$

(2.23)

Thus, the growth rate of per capita/worker output in steady-state does not depend on R&D intensity, as is the case in Schumpeterian growth theory, but on the growth rate of the population or the growth rate the growth rate of research input.

Moreover, we see by inserting the expression for A_t from equation (2.14) into (2.2) that, out of steady state, per worker output is given by

$$y_t = \left(\frac{\delta}{g_A}\right)^{\frac{1}{(1-\alpha)(1-\phi)}} X_t^{\frac{\lambda}{(1-\alpha)(1-\phi)}} \left(\frac{K_t}{Y_t}\right)^{\frac{\alpha}{1-\alpha}} h_t$$

(2.24)

Since h_t and $\left(\frac{K_t}{Y_t}\right)$ must be constant along a balanced growth path, (2.24) reduces in steady state to

$$y_t = c\left(\frac{\delta}{g_A}\right)^{\frac{1}{(1-\alpha)(1-\phi)}} X_t^{\frac{\lambda}{(1-\alpha)(1-\phi)}}$$

(2.25)

or

$$y_t = c\left(\frac{\delta}{g_A}\right)^{\frac{1}{(1-\alpha)(1-\phi)}} L_t^{\frac{\lambda}{(1-\alpha)(1-\phi)}}$$

(2.26)

where $c \equiv h\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}}$.

2.4. The hybrid semi-endogenous model

The hybrid semi-endogenous model can be derived by assuming a sectoral knowledge function of the type $\dot{A}_t = \delta A_t^{\phi} X_t^{\lambda}$ with $\phi < 1$ and assuming further that the number of sectors grows less than proportionally with population ($Q = L^{\beta}$, $\beta < 1$) (see, e.g., Jones, 1999). If $0 < \beta < 1$ and $\phi < 1$, then, after setting $Q = L^{\beta}$, equation (2.4) can be written as

$$\frac{\dot{A}_t}{A_t} = \delta \left(\frac{A_t}{X_t^{\frac{\lambda}{1-\phi}} L_t^{\frac{-\lambda\beta}{1-\phi}}} \right)^{\phi-1} \tag{2.27}$$

or

$$\frac{\dot{A}_t}{A_t} = \delta \left(\frac{X_t}{L_t} \right)^{\lambda} \left(\frac{A_t}{L_t^{\frac{\lambda(1-\beta)}{1-\phi}}} \right)^{\phi-1} \tag{2.28}$$

The latter equation shows that the hybrid model is also consistent with a constant R&D intensity and constant growth rate of knowledge, provided that $L_t^{\frac{\lambda(1-\beta)}{1-\phi}}$ is proportional to A_t. If the growth rate of knowledge is constant, equations (2.28) and (2.29) can be solved for A_t, giving

$$A_t = \left(\frac{\delta}{g_A} \right)^{\frac{1}{1-\phi}} X_t^{\frac{\lambda}{1-\phi}} L_t^{\frac{\lambda(-\beta)}{1-\phi}} \tag{2.29}$$

or

$$A_t = \left(\frac{\delta}{g_A} \right)^{\frac{1}{1-\phi}} \left(\frac{X_t}{L_t} \right)^{\frac{\lambda}{1-\phi}} L_t^{\frac{\lambda(1-\beta)}{1-\phi}} \tag{2.30}$$

Equation (2.29) implies that a constant growth rate of knowledge is also compatible with a situation in which $X_t^{\frac{\lambda}{1-\phi}} L_t^{\frac{-\lambda\beta}{1-\phi}}$ (and not only $X_t^{\frac{\lambda}{1-\phi}}$) is proportional to A_t. The constant growth rate of the knowledge stock is given by

$$g_A = \frac{\lambda}{1-\phi} g_X + \frac{\lambda(-\beta)}{1-\phi} g_L \tag{2.31}$$

In addition, it follows from (2.30) that the growth rate of knowledge can also be constant for a certain period of time even if R&D intensity is not constant, provided that the term $(X_t/L_t)^{\frac{\lambda}{1-\phi}} L_t^{\frac{\lambda(1-\beta)}{1-\phi}}$ is proportional to A_t. The previous "disequilibrium equation" (2.20) can thus be written as

$$g_A = \frac{\lambda}{1-\phi} g_{(X/L)} + \frac{\lambda(1-\beta)}{1-\phi} g_L \tag{2.32}$$

In steady state, this equation reduces to

$$g_A = \frac{\lambda(1-\beta)}{1-\phi} g_L \tag{2.33}$$

Now, inserting equation (2.31) into equation (2.11), and taking into account that the first term of equation (2.11) becomes $(\theta - 1)\beta g_L$ if $Q = L^\beta$, the steady-state growth rate of per capita output is given by

$$g_y = \frac{\lambda}{(1-\alpha)(1-\phi)} g_X + \left((\theta - 1)\beta + \frac{\lambda(-\beta)}{(1-\alpha)(1-\phi)}\right) g_L \tag{2.34}$$

Setting $g_X = g_L$ in (2.34) gives

$$g_y = \left((\theta - 1)\beta + \frac{\lambda(1-\beta)}{(1-\alpha)(1-\phi)}\right) g_L \tag{2.35}$$

Thus, in contrast to Schumpeterian growth models, the hybrid model implies a positive effect of population growth on output per capita growth even if there is no variety effect (if $\theta = 1$). If there is a variety effect (if $\theta > 1$), then it is even possible that the economic growth effect of population growth in the hybrid model is larger than the corresponding effect in (pure) semi-endogenous models.

3. Empirical literature

3.1. General remarks

This section reviews the literature testing the predictions of semi-endogenous models against the predictions of Schumpeterian growth models. In this literature three different approaches can be distinguished, which we denote here as Types I to III.

The Type I approach is based on the estimation of empirical specifications of equation (2.3), with the log of $\dot{A}$ as the dependent variable. Studies using this approach generally use the number of patent applications or patents granted as a measure of $\dot{A}_t$ and the patent stock or (real) GDP as a measure of A_t. The Type II approach is based on logarithmic specifications of equation (2.4), in which the dependent variable is $\log\dot{A}_t/A_t$. Studies based on this approach generally employ $\Delta\log A_{it}$ as an approximation for $\log\dot{A}_t/A_t$ and then use the differenced log of TFP as a measure of $\Delta\log A_{it}$ and the level of (the log of) TFP as a measure of (the log of) A_t. In addition, there are some studies that use equations (2.11), (2.22), and (2.24) (or variants of them) to discriminate between the theories. In these studies, the dependent variable is the growth rate and/or the level of output per capita (or per worker). We categorize these studies as Type III. Finally, this review includes one study that does not fit well into these categories. We denote the approach of this study as "unclassified".

Table 1 lists the studies included in our review and presents information on each study regarding the type of approach, the main variables of interest for our review, the type of analysis (time series analysis, panel analysis), the sample composition (number of countries/industries/firms and observation period), the econometric methods, and the evidence. As can be inferred from the table, our review includes 17 studies, three of which use the Type III approach, one is unclassified, two use both Type I and Type II, and eleven use either Type I or Type II. Type I, like Type II, is used in seven studies.

In the following subsections, we discuss these types in detail. Each subsection starts with a discussion of the specifications from the different approaches, followed by a detailed review of the findings from these approaches, including a critique of the existing research. The unclassified study is discussed in a separate subsection.

3.2. Type I approach

3.2.1. Specifications

Taking logs of equation (2.3) and adding an error term gives the following general empirical model:

$$\log\dot{A}_{it} = c_i + \phi\log A_{it} + \lambda\log X_{it} + \lambda(-\beta)\log Q_{it} + \varepsilon_{it} \tag{3.1}$$

where i denotes the cross-section unit and $c_i \equiv \log\delta_i$. If $\beta = 1$, equation (3.1) can be written as

$$\log\dot{A}_{it} = c_i + \phi\log A_{it} + \lambda\log\frac{X_{it}}{Q_{it}} + \varepsilon_{it} \tag{3.2}$$

If $\beta = 0$, equation (3.1) reduces to

$$\log\dot{A}_{it} = c_i + \phi\log A_{it} + \lambda\log X_{it} + \varepsilon_{it} \tag{3.3}$$

Writing the general form of equations (3.2) and (3.3) as

$$\log\dot{A}_{it} = c_i + \phi\log A_{it} + (\lambda\text{-}\lambda\beta)\log X_{it} + \lambda\beta\log\frac{X_{it}}{Q_{it}} + \varepsilon_{it} \tag{3.4}$$

we can see that if the estimated coefficient on $\log X_{it}$ in (3.2) is significantly positive while the estimated coefficient on $\log\frac{X_{it}}{Q_{it}}$ in (3.3) is not significantly different from zero, then this can be interpreted as evidence that $\beta = 0$, as *semi-endogenous growth models* predict; if the estimated coefficient on $\log X_{it}$ in (3.2) is insignificant while the estimated coefficient on $\log\frac{X_{it}}{Q_{it}}$ in (3.3) is positive and significant, then this can be interpreted as evidence that $\beta = 1$, as *Schumpeterian growth models* predict; finally, if both coefficients are significantly positive, then this can be interpreted as evidence in favor of $0 < \beta < 1$, as the *hybrid semi-endogenous model* predicts.

3.2.2. Findings

Based on equation (3.1), Furman et al. (2002) find that the estimates of ϕ are much smaller than one in some specifications. Another of their findings is that the estimated coefficient on $\log Q_{it}$ is always significantly negative, whereas the estimated coefficient on $\log X_{it}$ is always significantly positive. Given that the coefficient on $\log Q_{it}$ tends to be smaller in absolute value than the coefficient on $\log X_{it}$, their results imply $0 < \beta < 1$, which, together with the finding that ϕ tends to be smaller than one, is consistent with the "hybrid" semi-endogenous model.

Venturini (2012a) finds, also using equation (3.1), that the estimated coefficients on $\log X_{it}$ and $\log Q_{it}$ are significant and have the expected (positive and negative) signs in most cases.

Another result of his study is that although $\hat{\phi}$ is statistically significant and close to (but smaller than) one in most regressions, $\hat{\phi} = 1$ is rejected in the majority of cases. Similarly, his Wald tests reject both the null hypotheses that $\hat{\beta} = 0$ and the null hypothesis that $\hat{\beta} = 1$ in the majority of cases.

Luintel and Khan (2009) use a dynamic heterogeneous panel model to estimate the short-run elasticities of $\dot{A}_{it}$ with respect to A_{it}, X_{it}, and $\frac{X_{it}}{Q_{it}}$, both for each country in their sample and for their sample as a whole. From these short-run elasticities, they also calculate the country-specific long-run elasticities of the knowledge production function as given by equation (3.4). Their results suggest that the short- and long-run elasticities differ across countries, but the estimated effects of A_{it}, X_{it}, and $\frac{X_{it}}{Q_{it}}$ on $\dot{A}_{it}$ are significant in almost all countries. Their results also indicate that in almost all countries the long-run coefficient on $\log A_{it}$ is smaller than one and that the long-run coefficients on both $\log X_{it}$ and $\log \frac{X_{it}}{Q_{it}}$ are positive (suggesting $0 < \beta < 1$).

These results in favor of the hybrid semi-endogenous model are in contrast to those of Ang and Madsen (2011), Ang and Madsen (2015), Hu and Mathews (2005), and Fedderke and Liu (2017). Based on equations (3.2) and (3.3), Ang and Madsen (2011) find in all specifications that $\hat{\phi}$ is very close to one and that $\hat{\phi} = 1$ is not rejected. They also find that while the coefficients on $\log X_{it}$ are insignificant or have the „wrong" (negative) sign, the coefficients on $\log \frac{X_{it}}{Q_{it}}$ are positive and significant in all regressions (suggesting $\beta = 1$).

Support for Schumpeterian growth is also provided by Ang and Madsen (2015). Based on equations (3.1) and (3.3), they find that the estimated coefficient on $\log A_{it}$ is highly significant and close to one. Their results also show that the coefficient on $\log X_{it}$ is insignificant in equation (3.3). In equation (3.1), in contrast, the estimated λ coefficient is always positive and statistically significant, whereas the coefficient on $\log Q_{it}$ is always negative and significant. In addition, the coefficients on $\log X_{it}$ are similar or equal in absolute value to the coefficients on $\log Q_{it}$ in many regressions (suggesting $\beta = 1$).

In contrast, the results of Hu and Mathews (2005) are inconsistent with Schumpeterian growth, semi-endogenous growth, and hybrid semi-endogenous growth. The authors find that the estimated coefficient on $\log A_{it}$ is always significant and always greater than one. They also find, using equation (3.3), that the coefficient on $\log Q_{it}$ is negative and significant in all regressions, that the coefficient on $\log X_{it}$ is positive and significant in most regressions, and that the coefficient on $\log Q_{it}$ is always much greater in absolute value than that on $\log X_{it}$ (implying $\beta > 1$).

Finally, Fedderke and Liu (2017) provide time series estimates of equation (3.1) using (unlike the above studies) TFP as measure of A and the change in TFP as a measure of $\dot{A}$. They report that their results are "mixed"; in fact, however, their results are almost all inconsistent with both Schumpeterian growth and (hybrid) semi-endogenous growth, either because (*i*) the coefficients on logA_{it} are greater than one, (*ii*) the coefficients on logX_{it} are greater than one, (*iii*) the coefficients on logX_{it} are negative or insignificant, (*iv*) the coefficients on logQ_{it} are positive, (*v*) the coefficients on logQ_{it} are greater in absolute value than the coefficients on logX_{it}, or (*iv*) all coefficients are insignificant.

In closing this section, we note that the above studies yield mixed results on the size of the λ parameter. Venturini (2012a) and Ang and Madsen (2011, 2015) report λ estimates that are smaller, and sometimes much smaller, than 0.1. In contrast, the other studies reviewed here obtain estimates of λ that tend to be much greater than 0.1, and sometimes even greater than one.

3.2.3. Critique

A conceptual problem is that in many countries, especially the United States, the number of patents has grown much faster than GDP per worker, at least since the 1980s. Calculating the average growth rate of the stock of patent applications for the United States during the period 1980-2014, we obtain a value of 0.028. Using this value for g_A in equation (2.21), we would predict an average annual growth of per worker output through growth in the stock of knowledge of 0.043 (based on $\alpha \equiv 1/3$), but the actual average annual growth rate of output per worker over the period 1980-2014 was 0.016.[3] It thus appears that the number/stock of patents is not a good proxy for the flow/stock of ideas.[4] We come back to this point below.

A methodological problem is the potential non-stationarity of the data. If the dependent and explanatory variables in the above equations are stationary, then these equations can be used to estimate the desired relationships. If, however, the variables in equations (3.1)-(3.4) are non-stationary, then it has to be determined if they form a cointegrating set or not. If the variables in the above equations are non-stationary and not cointegrated, then regressions involving these variables may be spurious in the sense that they may indicate the existence of a significant relationship

[3] Following common practice, we constructed the patent stock from the number of patent applications using the perpetual inventory equation, $A_{it}=\dot{A}_{i0}+(1-\delta)A_{it\text{-}1}$, where δ is the depreciation rate. The data on the number of (domestic) patent applications were obtained from the World Intellectual Property Organization (WIPO) database (available at http://www.wipo.int/ipstats/en/). Consistent with the literature, the initial value of the patent stock was set equal to $A_{i0}=\dot{A}_{i0}/(g+\delta)$, where $\dot{A}_{i0}$ is the number of patent applications in the first year it is available, and g is average growth rate of the patent series between the first year with available data (1933) and the last year of the observation period. Following the literature, we used a depreciation rate of δ = 15%. Our data on GDP per worker are from the Penn World Tables 9.0 (available at https://www.rug.nl/ggdc/productivity/pwt/).

[4] Similarly, Jones (2016, p.20) argues that "[i]f each patent raises GDP by a constant percent, then a constant flow of new patents can generate a constant rate of economic growth. The problem with this approach (or perhaps the problem with the patent data) is that it breaks down after 1980 or so. Since 1980, the number of patents has risen by more than a factor of four, while growth rates are more or less stable."

between the variables when, in fact, the variables are unrelated.[5] If the variables in the above equations are non-stationary and cointegrated, inference from standard methods may be biased as well; however, this bias may result in a type II error and thus in failure to find a relationship that exists (see, e.g., Kao and Chiang 2000). The solution to the potential problem of non-stationarity is to pre-test the stationarity and cointegration properties of the data. If the variables are non-stationary and cointegrated, then an estimator should be used that provides valid inference in the presence of non-stationary data. If the variables are non-stationary but not cointegrated, the above equations should be estimated in first differences. However, of the above studies, only Venturini (2012a) and Fedderke and Liu (2017) test all variables in their estimating equation for non-stationarity and cointegration (in their online appendices), and since they find that their data are non-stationary and cointegrated, they use a cointegration estimator to estimate equation (3.1). Ang and Madsen (2015) do not test for cointegration, but they use levels equations, first difference equations, and a panel vector error model (VECM) to account for potential non-stationarity (and cointegration).[6] The other studies do not address possible problems due to potential non-stationarity.

Another related methodological problem is the order of integration. The order of integration is the number of times a variable must be differenced to achieve stationarity. In the case of more than two variables which are of a different order of integration and the order of the dependent variable is one ($I(1)$), while the explanatory variables are a mixture of $I(1)$ and $I(2)$ variables, these variables can cointegrate to a stationary (or $I(0)$) process if (*i*) there are at least two $I(2)$ variables, (*ii*) these $I(2)$ variables cointegrate to an $I(1)$ process, and (*iii*) this linear combination is then cointegrated with the other $I(1)$ variables.[7] Thus, if there is only one $I(2)$ regressor, then it cannot be cointegrated with the dependent variable. While this is intuitively obvious, the time-series econometrics literature suggests that standard cointegration tests and estimators, which assume that all variables are $I(1)$, can produce misleading test and estimation results when applied to unbalanced regressions with a mixture of $I(1)$ and $I(2)$ variables (see, e.g., Haldrup, 1998).[8] Before conducting cointegration tests and estimating the parameters of interest, it is therefore important to determine the order of integration of the variables. Furman et al. (2002), Hu and Mathews (2005), Luintel and

[5] The phenomenon of spurious regression is well known from the time-series econometrics literature. Kao (1999) shows that the tendency for spuriously indicating a relationship may even be stronger in panel data regressions than in pure time-series regressions.

[6] Ang and Madsen (2015, p. 85) note that "[they] have not used cointegration regression techniques in the baseline regressions because the dependent variable [the log of the patent stock] is found to be stationary." They (p. 85) also note that "to ensure that the estimates are not biased because of the conclusion of our unit root tests, the panel vector error-correction and first-difference estimators are also used in the robustness section under the assumption that the dependent variable contains a unit root."

[7] The $I(2)$ variables may also cointegrate directly to a stationary variable and thus form a cointegrating relationship with the $I(1)$ variables, provided that the $I(1)$ variables are cointegrated.

[8] False inference can also occur when the dependent variable is $I(0)$ but the regressor is $I(1)$, as we discuss in more detail in Section 3.3.3.

Khan (2009), however, do not examine the time-series properties of their data. Fedderke and Liu (2017) find that their variables are $I(1)$ (and cointegrated). Venturini (2012a) tests the variables for non-stationarity (and finds non-stationarity), but he does not determine the order of integration of the variables and thus does not account for the possibility of a mixture of $I(1)$ and $I(2)$ variables. Ang and Madsen (2011) investigate the time series properties of $\log X_{it}$, $\log Q_{it}$, and $\log \frac{X_{it}}{Q_{it}}$ (and report panel unit root test results that tend to support the stationarity of these variables), but they do not determine the order of integration of the log of patents and the log of the patent stock. Ang and Madsen (2015) find, somewhat surprisingly, that both the log of patent applications and the log of the patent stock are $I(0)$,[9] while $\log X_{it}$ and $\log Q_{it}$ are $I(1)$. Finally, Tables A1 and A2 in the Appendix provide results of time-series unit root tests for the United States and results of panel unit root tests for our panel of 19 countries for the levels and differences of the log of patent applications and the log of the patent stock. These tests suggest that the log of patent applications is $I(1)$, whereas the log of the patent stock is $I(2)$ (implying that the growth rate of the patent stock is $I(1)$). Besides the above mentioned methodological problems associated with the presence of $I(2)$ variables, this result once again implies that the stock of patents is a poor proxy for the stock of knowledge because (given that the growth rate of GDP per worker is generally found to be stationary) it is inconsistent with the idea that long-run growth in per capita income is driven by growth in the patent stock as a proxy for the knowledge stock.

As a final remark, we note that there are numerous studies estimating knowledge production functions without testing semi-endogenous growth models against Schumpeterian growth models. The focus of these studies is on testing either semi-endogenous growth theory (see, e.g. Abdih and Joutz, 2006) or Schumpeterian growth theory (see, e.g., Ulku, 2007), or they focus on examining the innovation impact of other variables that are not part of these theories, such as financial market development (see, e.g., Ang and Madsen, 2016). In general, these studies include either a measure of research input (and find a significantly positive coefficient on $\log X_{it}$) (see, e.g., Abdih and Joutz, 2006; Ang, 2010a; Luintel and Khan, 2017) or a measure of R&D intensity (and find significantly positive coefficient on $\log \frac{X_{it}}{Q_{it}}$) (see, e.g., Ulku, 2007; Ang, 2010b; Ang, 2014; Ang and Madsen, 2016). Thus, these studies are unable to detect partial product proliferation effects (and hence hybrid semi-endogenous growth) because they either do not allow at all for product proliferation effects or assume complete product proliferation.

In addition, by rewriting equation (3.1) as

[9] This would imply that despite the rise in income per capita over their sample period (1870-2010), the patent stock as a proxy for the knowledge stock stayed more or less the same and thus cannot have been the cause of growth during this period.

$$\log\dot{A}_{it} = c_i + \phi\log A_{it} + \lambda\log\frac{X_{it}}{Q_{it}} + \lambda(1\text{-}\beta)\log Q_{it} + \varepsilon_{it}$$

(3.5)

it becomes clear that a significant positive coefficient on $\log\frac{X_{it}}{Q_{it}}$ is not sufficient to reject semi-endogenous growth. If the case $\phi < 1$ cannot be ruled out with certainty and if $\log Q_{it}$ is significantly positively related to $\log\dot{A}_{it}$ (although omitted in many studies), it is still possible that (hybrid) semi-endogenous growth theory is valid even if the coefficient on $\log\frac{X_{it}}{Q_{it}}$ is positive and significant. In fact, Neves and Sequeira (2018) find in a meta-analysis of the effect of the stock of knowledge on the flow of knowledge, proxied by the number of patents, that the average estimate of ϕ is smaller than (but close to) one.

3.3. Type II approach

3.3.1. Specifications

The specific form of equation (2.4) for the case of *Schumpeterian growth* is equation (2.5), which after taking logs and approximating $\log\frac{\dot{A}_t}{A_t}$ by $\Delta\log A_t$ can be written in empirical form for the *i*th cross-sectional unit as

$$\Delta\log A_{it} = c_i + \lambda_i\log\frac{X_{it}}{Q_{it}} + \varepsilon_{it}$$

(3.6)

or

$$\Delta\log A_{it} = c_i + \lambda_i\left[\log X_{it} - \log Q_{it}\right] + \varepsilon_{it}$$

(3.7)

where $c_i \equiv \delta_i$. If $\Delta\log A_{it}$ is stationary it follows from equation (3.6) that, for a statistically meaningful relationship between $\Delta\log A_{it}$ and $\log\frac{X_{it}}{Q_{it}}$ to hold, $\log\frac{X_{it}}{Q_{it}}$ must also be stationary; in this case, the model can be estimated using conventional techniques. If, however, $\log X_{it}$ and $\log Q_{it}$ are non-stationary, then, as implied by equation (3.7), they must be cointegrated with a cointegrating vector (1, −1), which means that the estimated value of β_i in the regression

$$\log X_{it} = c_i + \beta_i\log Q_{it} + e_{it}$$

(3.8)

should be (close to) one and that the residuals from this regression should be stationary. Cointegration with a vector (1, −1) in turn implies the stationarity of the variable $\log\frac{X_{it}}{Q_{it}}$, whose coefficient should be significantly positive in equation (3.6) for Schumpeterian growth theory to be valid (see, e.g., Ang and Madsen, 2011). Combining equation (3.6) with the steady-state equation (2.15) from the semi-endogenous theory yields the following empirical model that nests both theories (see, e.g., Madsen, 2008):

$$\Delta\log A_{it} = c_i + \rho_{1i}\log\frac{X_{it}}{Q_{it}} + \rho_{2i}\Delta\log X_{it} + \varepsilon_{it} \tag{3.9}$$

Schumpeterian growth models predict ρ_{1i} ($\equiv \lambda_i$) > 0, whereas (hybrid) semi-endogenous growth models predict $\rho_{1i} = 0$ and ρ_{2i} ($\equiv \frac{\lambda_i}{1-\phi_i}$ in equation (2.15) or $\frac{\lambda_i(1-\beta_i)}{1-\phi_i}$ in equation (2.32)) > 0.

For the case of *semi-endogenous growth*, equation (2.13) is a specific form of equation (2.4) that can be written in the empirical form

$$\Delta\log A_{it} = c_i + (\phi_i - 1)\log A_{it} + \lambda_i\log X_{it} + \varepsilon_{it} \tag{3.10}$$

If the variables in this model are $I(0)$, then this model can be estimated directly with conventional techniques. If, however, $\Delta\log A_{it}$ is $I(0)$, whereas $\log A_{it}$ and $\log X_{it}$ are $I(1)$ (and not cointegrated), then equation (3.10) represents an unbalanced regression, which can produce spurious results (see, e.g., Stewart, 2011; Bekiros et al. 2018). To avoid such a mixture of $I(0)$ and $I(1)$ variables, it is useful to rewrite the above equation in the form

$$\Delta\log A_{it} = c_i - (1 - \phi_i)\left[\log A_{it} - \frac{\lambda_i}{1-\phi_i}\log X_{it}\right] + \varepsilon_{it} \tag{3.11}$$

or

$$\Delta\log A_{it} = c_i + \lambda_i\left[\log X_{it} - \frac{1-\phi_i}{\lambda_i}\log A_{it}\right] + \varepsilon_{it} \tag{3.12}$$

From equations (3.11) and (3.12) it can be concluded that if $\Delta\log A_{it}$ is stationary and $\log A_{it}$ and $\log X_{it}$ are non-stationary, the two non-stationary variables must be cointegrated with a cointegrating vector $(1, -\frac{\lambda_i}{1-\phi_i})$ or $(1, -\frac{1-\phi_i}{\lambda_i})$ (see, e.g., Ha and Howitt, 2007). The cointegrating relationship between $\log A_{it}$ and $\log X_{it}$ can thus be expressed as

$$\log A_{it} = c_i + \frac{\lambda_i}{1-\phi_i}\log X_{it} + e_{it} \tag{3.12}$$

Since equation (3.12) is an empirical version of equation (2.14), the term $\frac{\lambda_i}{1-\phi_i}$ can be interpreted as the long-run, or steady state, elasticity of knowledge with respect to research input. If the degree of returns to scale ϕ is less than unity, then this elasticity should be significantly positive.

Alternatively, the cointegrating relationship between the two variables can be expressed as

$$\log X_{it} = c_i + \frac{1-\phi_i}{\lambda_i}\log A_{it} + e_{it} \tag{3.13}$$

In this reverse regression, the coefficient on $\log A_{it}$ should be significantly positive.

In the case of the *hybrid semi-endogenous model*, equation (2.4) can be written as equation (2.27). An empirical form of this equation is

$$\Delta\log A_{it} = c_i - (1-\phi_i)\left[\log A_{it} - \frac{\lambda_i}{1-\phi_i}\log X_{it} + \frac{\lambda_i\beta_i}{1-\phi_i}\log Q_{it}\right] + \varepsilon_{it} \tag{3.14}$$

which can be rewritten as

$$\Delta\log A_{it} = c_i + \lambda_i\left[\log X_{it} - \lambda_i\beta_i\log Q_{it} - \frac{1-\phi_i}{\lambda_i}\log A_{it}\right] + \varepsilon_{it} \tag{3.15}$$

It follows from these equations that if $\Delta\log A_{it}$ is stationary and $\log A_{it}$, $\log X_{it}$, and $\log Q_{it}$ are non-stationary, then $\log A_{it}$, $\log X_{it}$, and $\log Q_{it}$ must be cointegrated for the hybrid semi-endogenous model to hold. In the case of one cointegrating vector among the three non-stationary variables, the long-run relationship can be written as

$$\log A_{it} = c_i + \frac{\lambda_i}{1-\phi_i}\log X_{it} - \frac{\lambda_i\beta_i}{1-\phi_i}\log Q_{it} + e_{it} \tag{3.16}$$

This equation, which is an empirical version of equation (2.29), implies that the hybrid semi-endogenous model cannot be rejected if the coefficient $\log X_{it}$ is significantly positive and the coefficient on $\log Q_{it}$ is significantly negative.

The hybrid model is also supported if, alternatively, the coefficients on $\log Q_{it}$ and $\log A_{it}$ in the reverse cointegrating regression

$$\log X_{it} = c_i + \lambda_i\beta_i\log Q_{it} + \frac{1-\phi_i}{\lambda_i}\log A_{it} + e_{it} \tag{3.17}$$

are significant and positive.

Finally, the following two equations are also empirical versions of equation (2.4):

$$\Delta\log A_{it} = c_i - (1-\phi_i)\left[\log A_{it} - \frac{1}{1-\phi_i}\log X_{it}\right] + (\lambda_i - 1)\left[\log X_{it} - \frac{\lambda_i\beta_i}{\lambda_i - 1}\log Q_{it}\right] + \varepsilon_{it} \tag{3.18}$$

$$\Delta\log A_{it} = c_i - (1-\phi_i)\left[\log A_{it} - \frac{\lambda_i(1-\beta_i)}{1-\phi_i}\log Q_{it}\right] + \lambda_i\log\frac{X_{it}}{Q_{it}} + \varepsilon_{it} \tag{3.19}$$

Equation (3.18) implies that the hybrid semi-endogenous model may also be valid if there are two cointegrating relationships between the level variables—that is, (*i*) if $\log A_{it}$ and $\log X_{it}$ are cointegrated with a cointegrating vector $(1, -\frac{1}{1-\phi_i})$ in which the second element is negative and (*ii*) if $\log X_{it}$ and $\log Q_{it}$ are cointegrated with a cointegrating vector $(1, -\frac{\lambda_i\beta_i}{\lambda_i-1})$ in which the second element is positive. If the second element of the cointegrating vector $(1, -\frac{\lambda_i\beta_i}{\lambda_i-1})$ is positive, then $\log\frac{X_{it}}{Q_{it}}$ is non-stationary (because the stationarity of $\log\frac{X_{it}}{Q_{it}}$ implies a cointegrating vector $(1, -1)$ in which the second element is equal to –1). If, in contrast, $\log X_{it}$ and $\log Q_{it}$ are not cointegrated, then $\log\frac{X_{it}}{Q_{it}}$ is also non-stationary (as in the case where $\log X_{it}$ and $\log Q_{it}$ are cointegrated with a cointegrating vector different from $(1, -1)$) and equation (3.18) reduces to equation (3.11). Thus, non-stationarity of $\log\frac{X_{it}}{Q_{it}}$ is consistent with both the hybrid semi-endogenous growth model and "pure" semi-endogenous growth models, but the latter implies (in the case of a non-stationary R&D intensity variable) that $\log X_{it}$ is not cointegrated with $\log Q_{it}$. It therefore follows from the invariance of cointegration relationships to model extensions that if $\log A_{it}$ and $\log X_{it}$ are cointegrated, while $\log X_{it}$ and $\log Q_{it}$ are not cointegrated, the hybrid semi-endogenous growth model can be rejected.[10]

Non-stationarity of $\log\frac{X_{it}}{Q_{it}}$ is inconsistent and stationarity of $\log\frac{X_{it}}{Q_{it}}$ is consistent with Schumpeterian growth models if $\Delta\log A_{it}$ is stationary, as argued above. However, as implied by equation (3.19), and already discussed in Sections 2.3 and 2.4, (hybrid) semi-endogenous growth models may also be valid if R&D intensity is stationary. The necessary condition is that $\log A_{it}$ is cointegrated with $\log Q_{it}$—with a cointegrating vector $(1, -\frac{\lambda_i(1-\beta_i)}{1-\phi_i})$ in the case of hybrid semi-endogenous growth and a cointegrating vector $(1, -\frac{\lambda_i}{1-\phi_i})$ in the case of semi-endogenous growth. Thus, the validity of (hybrid) semi-endogenous growth models requires cointegration between $\log A_{it}$ and $\log X_{it}$ (and $\log X_{it}$ and $\log Q_{it}$) or, if $\log\frac{X_{it}}{Q_{it}}$ is stationary, cointegration between $\log A_{it}$ and $\log Q_{it}$, whereas the validity of the Schumpeterian theory requires not only that $\log\frac{X_{it}}{Q_{it}}$ is stationary, but also that $\log A_{it}$ is not cointegrated with $\log Q_{it}$ and hence with $\log X_{it}$.[11] The implication is that (hybrid) semi-endogenous growth models can be rejected in favor of Schumpeterian growth models if, and only if, $\log X_{it}$ is cointegrated with $\log Q_{it}$ but not cointegrated with $\log A_{it}$.

In addition, equation (3.19) (which is an empirical version of equation (2.28)) shows that (hybrid) semi-endogenous growth models do not predict that the coefficient on $\log\frac{X_{it}}{Q_{it}}$ in a regression of $\Delta\log A_{it}$ on $\log\frac{X_{it}}{Q_{it}}$ will be zero. Although Schumpeterian growth models predict that this coefficient will be significantly positive, a significant positive coefficient on $\log\frac{X_{it}}{Q_{it}}$ in equation (3.6) does not imply a rejection of (hybrid) semi-endogenous growth. The reason why (hybrid) semi-endogenous growth models are consistent with a significant positive coefficient estimate for R&D intensity is that these models suggest that if $\Delta\log A_{it}$ and $\log\frac{X_{it}}{Q_{it}}$ are stationary and population growth is constant, temporary increases in R&D intensity lead to temporary increases in the growth rate of research effort and thereby to temporary increases in the growth rate of A_{it}.

[10] The invariance of cointegration relationships to model extensions means that if there is cointegration between a set of variables, then this stationary relationship also exists in extended variable space.

[11] If X_{it} and Q_{it} behave so that their log ratio is stationary, then their logarithms are cointegrated (with a cointegrating vector $(1,-1)$). If the pairs ($\log X_{it}$, $\log Q_{it}$) and ($\log A_{it}$, $\log Q_{it}$) are cointegrated, it follows necessarily that the pair ($\log A_{it}$, $\log X_{it}$) is also cointegrated.

It follows from the last two paragraphs that studies testing only Schumpeterian theory—by testing the stationarity of $\log\frac{X_{it}}{Q_{it}}$ and running regressions of the form of equation (3.6)—are only able to reject Schumpeterian models (if $\log\frac{X_{it}}{Q_{it}}$ is non-stationary or $\log\frac{X_{it}}{Q_{it}}$ is stationary but the coefficient on $\log\frac{X_{it}}{Q_{it}}$ in equation (3.6) is insignificant), but these studies, which include Zachariadis (2003, 2004), among others, are unable to reject (hybrid) semi-endogenous models in favor of Schumpeterian models.

Similarly, it follows from the discussion of equations (3.17) and (3.18) that studies, such as those of Ang (2011a, 2011b) and Bottazzi and Peri (2007, 2015), which are based on a "pure" semi-endogenous model—and thus do not include $\log Q_{it}$—are unable to reject the hybrid in favor of the "pure" semi-endogenous growth theory.

The studies whose findings are summarized below do not suffer from these limitations because they use at least two of the above specifications to distinguish between the theories or they use equation (3.17), which includes both $\log A_{it}$ and $\log Q_{it}$.

3.3.2. Findings

With the exception of Venturini (2012b), all studies we discuss here measure A_{it} by TFP_{it} and $\Delta\log A_{it}$ by the differenced log of TFP_{it}. The common finding of these studies is that $\log A_{it}$ is $I(1)$, so that $\Delta\log A_{it}$ is $I(0)$ (in what follows the country index i is omitted whenever data for an aggregate time-series analysis of a single country are used).

The earliest of these studies is the work of Ha and Howitt (2007), who find that $\log\frac{X_t}{Q_t}$ is stationary, but they find no evidence of cointegration between $\log A_t$ and $\log X_t$. Similar findings in support of the Schumpeterian theory are reported by Madsen (2008), Saunoris and Payne (2011), Madsen et al. (2010a), and Ang and Madsen (2011). Madsen (2008) finds strong evidence that $\log X_{it}$ and $\log Q_{it}$ are cointegrated and that the coefficient on $\log Q_{it}$ in equation (3.8) is always positive and sometimes close to one. In contrast, there is relatively weak evidence of cointegration between $\log A_{it}$ and $\log X_{it}$. In addition, his results show that, although there is some evidence of cointegration between $\log A_{it}$, $\log X_{it}$, and $\log Q_{it}$, the estimated coefficient on $\log Q_{it}$ in equation (3.16) always has the "wrong" (positive) sign (to indicate hybrid semi-endogenous growth) and the estimated coefficient on $\log X_{it}$ is insignificant and/or has the "wrong" (negative) sign.

Similarly, Saunoris and Payne (2011) find that $\log X_t$, $\log Q_t$, and $\log A_t$ are $I(1)$ and cointegrated, but the coefficient on $\log A_t$ in equation (3.17) is always insignificant (at the 5% level) and has the "wrong" sign. In contrast, the coefficient on $\log Q_t$ in equation (3.17) is always significant, and the null hypothesis that the coefficient on logQ_t is equal to one cannot be rejected (suggesting $\beta = 1$ and $\lambda = 1$).

Madsen et al. (2010a) and Ang and Madsen (2011) find evidence of both stationarity of $\log\frac{X_{it}}{Q_{it}}$ and cointegration between logX_{it} and logQ_{it} (although the magnitudes of the estimated cointegrating vectors are, with one exception, always not as predicted by Schumpeterian growth models).[12] In contrast, there is little evidence of cointegration between logA_{it} and logX_{it} in the study by Ang and Madsen (2011), while Madsen et al. (2010a) find some evidence of cointegration between logA_{it} and logX_{it}, but the second element of the cointegrating vector has the "wrong" sign and/or is insignificant or implausibly large (in absolute value).

Finally, Madsen (2008) and Madsen et al. (2010a) also find that the coefficient on $\log\frac{X_{it}}{Q_{it}}$ in equation (3.9) is positive and significant in most regressions, whereas the coefficient on ΔlogX_{it} is insignificant or has the "wrong" sign in most regressions; and in the study by Ang and Madsen (2011), the coefficients on $\log\frac{X_{it}}{Q_{it}}$ in equation (3.9) are more often significant than the coefficients on ΔlogX_{it}.

These results are in contrast to those of Barcenilla-Visús et al. (2014), Fedderke and Liu (2017), and Venturini (2012b). Barcenilla-Visús et al. (2014) find that logA_{it}, logX_{it}, logQ_{it}, and $\log\frac{X_{it}}{Q_{it}}$ are $I(1)$. Their results also suggest that logA_{it} is cointegrated with logX_{it} and that logX_{it} is cointegrated with logQ_{it} (so that there are two cointegrating relationships between logA_{it}, logX_{it}, logQ_{it}, as described by equation (3.18)). Thus, although Barcenilla-Visús et al. (2014) claim that they find strong support for semi-endogenous growth, the results of their integration and cointegration tests are consistent with the hybrid semi-endogenous model. However, in contrast to what one would expect from the hybrid semi-endogenous model, their estimated coefficients on logX_{it} have mixed signs in the cointegrating regressions of logA_{it} on logX_{it}, and the estimated cointegration relationships between logX_{it} on logQ_{it} are either positive or insignificant (implying that, contrary to the predictions of the hybrid semi-endogenous model with two cointegrating relationships, the second element of the cointegrating vector between logX_{it} on logQ_{it} is not negative). Thus, their results are inconsistent with Schumpeterian growth, semi-endogenous growth, and hybrid semi-endogenous growth.

Fedderke and Liu (2017) (also) estimate equation (3.17). Similar to Barcenilla-Visús et al. (2014), their results are almost all inconsistent with the predictions of both theories, either because (*i*) the coefficients on logA_{it} are negative (implying $\phi > 1$), (*ii*) the coefficients on logQ_{it} are negative

[12] Madsen et al. (2010a) find that all of their (logged) measures of R&D intensity are stationary; the only exception is the ratio of researchers to total employment.

(implying $\beta < 0$ if $\lambda > 0$), or (*iii*) the coefficients on $\log Q_{it}$ are greater than one (implying $\beta > 1$ if $\lambda > 0$). In addition, they find, inconsistent with Schumpeterian growth, that while $\Delta\log A_{it}$ is $I(0)$, $\log \frac{X_{it}}{Q_{it}}$ is $I(1)$.

Finally, Venturini (2012b) estimates equations (3.6) and (3.10) using the patenting rate as the dependent variable; the number of innovating firms (normalized by output or employment) acts as proxy for A_{it} in equation (3.10).[13] He finds that the relationship between the patenting rate and R&D intensity is not significant when control variables are added to the equation. In contrast, the coefficient on $\log A_{it}$ is always significantly negative but smaller (in absolute terms) than one (suggesting $0 < \phi < 1$), while the coefficient on $\log X_{it}$ is always significantly positive but smaller than one (suggesting $0 < \lambda < 1$).

As a final remark, we note that while Madsen (2008) reports estimated values of λ that are positive, statistically significant but close to zero in most regressions (typically below 0.04), and while the λ estimates in Madsen et al. (2010a) and Venturini (2012b) lie between 0.1 and 0.9, Ang and Madsen (2011) report estimates of $\lambda > 1$.

3.3.3. Critique

The problem with the results of Venturini (2012b) and Fedderke and Liu (2017) is that they may be spurious due to non-stationarity and non-cointegration. Fedderke and Liu (2017) find that $\log X_{it}$, $\log Q_{it}$, and $\log A_{it}$ (measured by the log of TFP_{it}) are $I(1)$, but they do not test for cointegration.[14] Venturini (2012b) does not test his variables for stationarity, implying that they may be non-stationary and not cointegrated.[15] Moreover, the study by Venturini (2012b) is potentially limited by

[13] Venturini (2012b) uses a different notation in his equations, but his equations correspond to equations (3.6) and (3.10) (with the modification that he uses lagged explanatory variables).

[14] Fedderke and Liu (2017) report (in their online appendix) tests of the null hypothesis that there are no cointegrating relationships between $\log \dot{A}_{it}$, $\log A_{it}$, $\log X_{it}$, and $\log Q_{it}$ against the alternative that there is at least one cointegrating relationship between these four variables, but they do not test for cointegration between $\log X_{it}$, $\log Q_{it}$, and $\log A_{it}$. The point is that rejection of the null of no cointegrating vectors among the four variables does not imply cointegration among the three variables. More specifically, if $\log \dot{A}_{it}$, $\log A_{it}$, $\log X_{it}$, and $\log Q_{it}$ are cointegrated with one cointegrating vector, then there cannot be a cointegrating relationship between $\log X_{it}$, $\log Q_{it}$, and $\log A_{it}$.

[15] In order to achieve stationarity, Venturini (2012b) normalizes his right-hand-side variables in equation (3.10) by expressing them as ratios to output or employment. However, even such ratios may be non-stationary, and Venturini (2012b) does not provide stationarity tests for his (explanatory and dependent) variables. Thus, the possibility cannot be excluded that his variables are non-stationary processes. Specifically, the log of the patenting rate may be non-stationary when the log of the number of patent applications is $I(1)$ and the log of the patent stock is $I(2)$. In fact, we find evidence that this is the case, as discussed in the previous section. Moreover, the logarithmic patenting rate is an approximation of the growth rate of the stock of patents, and thus necessarily evolves roughly proportional to the first difference of the log of the patent stock. Thus, our finding that the first difference of the log of the patent stock has a unit root implies that the logarithmic patenting rate is also likely to be unit root process. To account for the possibility of spurious regressions due to unit roots in the data, Venturini (2012b) tests the residuals of his estimated equations for stationarity, at least in the working paper version of his article (Venturini, 2011). However, the panel unit root test he uses is a test for unit roots in univariate time series, and thus is not designed to test residuals from multivariate regressions for stationarity. In other words, the panel unit root test he applies to test the residuals of his regressions for stationarity is not a formal test for cointegration.

the use of the (logarithmic) patenting rate, which, as an approximation of the growth rate of the stock of patents, may be a poor proxy for the growth rate of the stock of knowledge for the reasons discussed in Section 3.2.3.

A general problem with some of the studies is that unit root tests can suffer from large size distortions that lead to spurious rejections of the unit root null hypothesis when the lag length is too small (see, e.g., Ng and Perron, 2001). This problem of spurious rejections may be exacerbated in panel data because panel unit root tests are known to have higher power to reject the null than unit root tests based on individual time series. The point here is that the validity of Schumpeterian growth theory depends on whether both $\Delta \log A_{it}$ and $\log \frac{X_{it}}{Q_{it}}$ are stationary and that studies such as those of Ha and Howitt (2007), Madsen et al. (2010a), and Ang and Madsen (2011), which provide formal evidence for the stationarity of both variables using unit root tests, may suffer from this problem if the lag length in the unit root tests is not adequately specified. In fact, visual inspection of the data in the studies of Ha and Howitt (2007), Madsen et al. (2010a), and Ang and Madsen (2011) shows that while TFP growth rates exhibit no large persistent changes, the measures of $\log \frac{X_{it}}{Q_{it}}$ tend to have an upward trend. Similarly, it can be seen from Figure 1, which graphs the log of the share of researchers in total employment and the growth rate of TFP for the United States and for the 19 countries in our study over the period 1980-2014, that this measure of $\log \frac{X_{it}}{Q_{it}}$ shows a very strong upward trend, whereas TFP growth exhibits no persistent increase. This casts doubt on the results of the unit root tests on the log levels of R&D intensity reported in the studies by Ha and Howitt (2007), Madsen et al. (2010a), and Ang and Madsen (2011) and thus on the evidence for Schumpeterian growth. [16]

[16] Jones (1995 (2002, p.226) graphs the share of researchers in total employment for the United States from 1950 to 1993 (based on data with interpolated values). The graph shows that this measure of R&D intensity has increased about three-fold during this period—and thus appears to be clearly non-stationary. Nevertheless, we find that the log of the share of researchers in total employment is stationary when we apply the ADF unit root test based on standard lag selection criteria (such as the Schwarz information criterion and the modified Schwarz information criterion) to the data from the graph (available at https://web.stanford.edu/~chadj/datasets.html). The point is that standard lag selection criteria select one lag. In contrast, the general-to-specific criterion of Hall (1994) selects four lags. In fact, when we use four lags, the ADF indicates that the share of researchers in total employment is non-stationary.

Another potential problem with these studies is that they use time series on R&D expenditures and/or the number of researchers with large, interpolated data gaps.[17] This can bias tests of cointegration in favor of the null of no cointegration between $\log A_{it}$ and $\log X_{it}$ and thus lead to erroneous conclusions regarding the rejection of semi-endogenous growth if there are large discrepancies between the interpolated and true values of these variables.[18]

The same problem may occur with the use R&D expenditures as a proxy for research effort: Since there is no R&D-specific deflator based on the wages of R&D workers and the prices of R&D investment goods, nominal R&D expenditures are deflated by an overall deflator, typically based on an aggregate wage index and/or an aggregate price deflator. This may lead to spurious fluctuations in R&D that reflect mismeasurement of the prices of R&D inputs rather than true changes in R&D activity (see, e.g., Barlevy, 2007). As a consequence, cointegration tests may not be able to detect cointegration between the log of TFP and the log of the true research effort.

In addition, the coefficient on $\Delta \log X_{it}$ in equation (3.9) may be biased toward zero if research effort is measured with error. The lack of significance of the coefficient on the growth rate of R&D expenditures/R&D labor therefore does not necessarily imply evidence against semi-endogenous growth but may be due instead to measurement error.

Another problem with equation (3.9) can be illustrated as follows. As noted above, the term $\frac{\lambda_i}{1-\phi_i}$ in equation (3.12) can be interpreted as the long-run, or steady state, elasticity of knowledge with respect to research input. If $\log A_{it}$ and $\log X_{it}$ are $I(1)$, then the two series must be cointegrated for a long-run equilibrium relationship, as given by equation (3.12), to exist. If the variables are cointegrated, $\frac{\lambda_i}{1-\phi_i}$ can be estimated directly from equation (3.12) (using a cointegration estimator) or indirectly from an autoregressive distributed lag model of the form

---

[17] Madsen et al. (2008) report in the working paper version of their 2010 article on "The Indian growth miracle and endogenous growth" that the data on the number of researchers between 1950 and 1990 are available only at ten-year intervals. Ang and Madsen (2011) report that data for China (Korea) [Singapore] {Taiwan} on R&D expenditures and the number of R&D workers are not available for the years 1953–1959 and 1961–1977 (1953–1967) [1953–1970] {1953–1970}. Ha and Howitt (2007) claim that their data on the number of scientists and engineers engaged in R&D for the period 1950–1997 are from the Science and Engineering Indicators 2000 published by the National Science Foundation (which can be obtained at https://wayback.archive-it.org/5902/20150627162232/http://www.nsf.gov/statistics/seind00/append/appa.htm in Appendix Table 3–25). However, these data are available from the Science and Engineering Indicators 2000 only for the period 1979–1997 and contain many gaps (at least for the United States). It should perhaps be noted here that the Industrial Research and Development Information System contains historical data on the number of scientists and engineers engaged in R&D for the United States starting in 1954 (available at https://wayback.archive-it.org/5902/20181004145214/https://www.nsf.gov/statistics/iris/search_hist.cfm?indx=24), but these data are also not available for many years (1955, 1956, 1959, 1960, 1961, 1962, 1965, 1966, 1969, 1970, and 1982). In addition, these data are available only for January rather than for the entire year (as annual average) during the period 1954–1981.

[18] Interpolation does not pose problems if the true values are cointegrated with the interpolated values, in which case the measurement error is stationary. It is well known that stationary measurement error has no serious effect on tests for cointegration, whereas in the case of non-stationary measurement error, cointegration tests fail to find cointegration between the variables involved.

$$\log A_{it} = c_{0i} + \phi_i \log A_{it\text{-}1} + \lambda_i \log X_{it} + \varepsilon_{it}$$

(3.20)

It is well known from the time series literature that equation (3.20) can be rewritten in error-correction form as

$$\Delta \log A_{it} = c_{1i} - (1-\phi_i)\left[\log A_{it\text{-}1} - \frac{\lambda_i}{1-\phi_i}\log X_{it\text{-}1} - c_i\right] + \lambda_i \Delta \log X_{it} + \varepsilon_{it}$$

(3.21)

where the term in brackets is the so-called error-correction term representing deviations from the long-run equilibrium relationship, and the term $-(1-\phi_i)$ represents the so-called error-correction coefficient measuring the speed of adjustment to the long-run equilibrium. According to the Granger Representation Theorem (Engle and Granger, 1987), the error-correction coefficient must be non-zero if $\log A_{it}$ and $\log X_{it}$ are cointegrated and long-run Granger causality runs from $\log X_{it}$ to $\log A_{it}$. In the case of semi-endogenous growth, the error-correction model given by equation (3.21) thus explains the growth rate of A_{it} by the growth in X_{it} and the past disequilibrium. The implication is that if the estimate of $-(1-\phi_i)$ is not significantly different from zero, then a long-run relationship between $\log A_{it}$ and $\log X_{it}$ does not exist, so that semi-endogenous growth can be rejected; if, however, the estimate of $-(1-\phi_i)$ is significantly negative, and semi-endogenous growth can therefore not be rejected, then simple models in growth rates may produce misleading results—not only because the error-correction term accounts for the long-run relationship between $\log A_{it}$ and $\log X_{it}$, but also because the omission of the error-correction term may bias the coefficient on $\Delta \log X_{it}$ down. The problem of equation (3.9) is thus that it is misspecified by the omission of the lagged error-correction term, which should be significant and negative if semi-endogenous growth theory is correct; an empirical growth model that nests both semi-endogenous and Schumpeterian growth should therefore not only include $\Delta \log X_{it}$ and $\log \frac{X_{it}}{Q_{it}}$ (provided that $\log \frac{X_{it}}{Q_{it}}$ is stationary), but also the error-correction term. However, to be fair, it should be noted again that Madsen (2008), Madsen et al. (2010a), and Ang and Madsen (2011), who use equation (3.9), find little evidence of cointegration between $\log A_{it}$ and $\log X_{it}$.

A final problem with the use of equation (3.9) is that it represents an unbalanced regression when the growth rate of TFP is stationary and R&D intensity is non-stationary,[19] in which case its use can lead to spurious results. As shown by Bekiros et al. (2018), the tendency for spuriously indicating a relationship in unbalanced regressions may even be stronger in panel than in time series

[19] Balance in such a situation requires the presence of an additional $I(1)$ regressor and cointegration between the $I(1)$ regressors.

data. The point is that in the studies that estimate equation (3.9) (or equation (3.6)) and report statistically significant coefficients on $\log\frac{X_{it}}{Q_{it}}$, such as Madsen (2008) and Madsen et al. (2010a), the significance is necessarily spurious if the regression is unbalanced since the true slope on the integrated regressor is necessarily zero in an unbalanced regression. However, we note here again that the results in Madsen (2008) and Madsen et al. (2010a) suggest that R&D intensity is stationary.

3.4. Type III approach

3.4.1. Specifications

An empirical form of the steady state equation (2.12) for the growth rate of per capita/worker output from *Schumpeterian growth theory* can be expressed as

$$g_{y,it} = c_i + b_1 g_{L,it} + b_2 \frac{X_{it}}{Q_{it}} + \varepsilon_{it} \tag{3.22}$$

As discussed in Section 2, *Schumpeterian growth theory* suggests that the coefficient on $g_{L,it}$ ($\equiv (\theta$-1)) may be positive or insignificant, whereas the coefficient on $\frac{X_{it}}{Q_{it}}$ ($\equiv \frac{\lambda\delta}{1-\alpha}$) should be positive and significant. In contrast, *semi-endogenous growth theory* predicts that the steady-state rate of per capita output growth is proportional to population growth (with a coefficient on $g_{L,it}$ equal to $\frac{\lambda}{(1-\alpha)(1-\phi)}$) and independent of R&D intensity. Minniti and Venturini's (2017) idea is therefore that equation (3.22) nests the steady state equation (2.22) for the growth rate of per capita/worker output from semi-endogenous growth theory, and thus can be used as a basis to test semi-endogenous growth models against Schumpeterian growth models. Based on this idea, $b_1 > 0$ and $b_2 = 0$ can be interpreted as evidence for semi-endogenous growth models, whereas $b_2 > 0$ can be interpreted as evidence for Schumpeterian growth models.

Another idea in the literature is to estimate the parameters b_1 and b_2 separately in two models (see, e.g., Laincz and Peretto, 2006),

$$g_{y,it} = c_i + b_1 g_{L,it} + \varepsilon_{it} \tag{3.23}$$

$$g_{y,it} = c_i + b_2 \frac{X_{it}}{Q_{it}} + \varepsilon_{it}$$

(3.24)

Replacing $g_{L,it}$ on the right hand side of equation (3.23) by $g_{X,it}$ yields an alternative empirical version of the steady state equation for the growth rate of per capita/worker output from *semi-endogenous growth theory* (based on equation (2.23)),

$$g_{y,it} = c_i + b_1 g_{X,it} + \varepsilon_{it}$$

(3.25)

Replacing $\frac{X_{it}}{Q_{it}}$ on the right hand side of equation (3.24) by $\frac{L_{it}}{Q_{it}}$ gives an alternative empirical form of the equation for the steady state growth rate from *Schumpeterian growth theory*,[20]

$$g_{y,it} = c_i + b_2 \frac{L_{it}}{Q_{it}} + \varepsilon_{it}$$

(3.26)

A third idea, suggested by Madsen et al. (2010b), combines equation (3.25) with equation (3.27),

$$g_{y,it} = c_i + b_3 \log \frac{X_{it}}{Q_{it}} + \varepsilon_{it}$$

(3.27)

to get

$$g_{y,it} = c_i + b_1 g_{X,it} + b_3 \log \frac{X_{it}}{Q_{it}} + \varepsilon_{it}$$

(3.28)

where b_3, like b_2, must be greater than one for Schumpeterian growth theory to hold (Madsen et al., 2010b).

Thus, a positive and significant b_2 or b_3 is a necessary condition for the validity of Schumpeterian growth theory. A necessary (but not sufficient) precondition is that (the log of) $\frac{X_{it}}{Q_{it}}$ (or $\frac{L_{it}}{Q_{it}}$) is stationary, or, equivalently, that (the log of) X_{it} (L_{it}) and (the log of) Q_{it} are cointegrated

[20] The intuition behind equation (3.26) is that with a constant share of labor devoted to R&D, the number of people employed in R&D is proportional to total employment, so that, in steady state, the ratio of total employment to the number of sectors or firms must also be constant. With a constant fraction of labor allocated to R&D, higher employment per sector or firm means more researchers per sector or firm and thus a higher growth rate in Schumpeterian growth models (see, e.g., Laincz and Peretto, 2006).

with a cointegrating vector (1, −1)) (see, e.g., Madsen et al., 2010b). The reason for this precondition is the prediction of Schumpeterian growth models that growth rates of GDP per capita/worker should be proportional to the (log-)level of R&D intensity, implying that if $g_{y,it}$ is stationary (as generally found in the literature), then (the log of) $\frac{X_{it}}{Q_{it}}$ ($\frac{L_{it}}{Q_{it}}$) must also be stationary.

As far as b_1 is concerned, it should be noted that this coefficient does not allow a distinction between pure semi-endogenous models and the hybrid semi-endogenous model in equations (3.22) and (3.23). As discussed in Section 2, the steady-state growth rate of output per capita/worker can be expressed as the product of the growth rate of the population and the term $\left((\theta-1)\beta+\frac{\lambda(1-\beta)}{(1-\alpha)(1-\phi)}\right)$ in the case of the hybrid model (see equation (2.35)), or as the product of the population growth rate and $\frac{\lambda}{(1-\alpha)(1-\phi)}$ in semi-endogenous theory (see equation (2.22)). A positive and significant value for b_1 in equations (3.22) and (3.23) may therefore represent both $\frac{\lambda}{(1-\alpha)(1-\phi)}$ and $\left((\theta-1)\beta+\frac{\lambda(1-\beta)}{(1-\alpha)(1-\phi)}\right)$ and is thus consistent with both pure and hybrid semi-endogenous models. Similarly, equations (3.25), and (3.28) unable to discriminate between these models because they do not include the growth rate of the population, and semi-endogenous growth models predict that the growth rate of output per capita/worker in the steady state depends only on the growth rate of research effort (and the parameters σ, λ, α, and ϕ) (see equation (2.23)), while the hybrid model predicts that the growth rate of output per capita/worker in the steady state depends on both the growth rate of research effort and the growth rate of the population (and the parameters λ, α, ϕ, β, and θ) (see equation (2.34)). In this sense, the implicit intention of equations (3.22)-(3.28) is to is to provide a test of Schumpeterian growth models against pure *and* hybrid semi-endogenous models.

Finally, Madsen et al. (2010b) test *semi-endogenous growth theory* (over a long historical period) using a variant of the per capita production function in equation (2.24) in which land T is included as a factor of production but human capital per worker is excluded. Taking logs of both sides of this equation and adding an error term yields

$$\log y_{it}= c_i+\frac{\lambda}{(1-\alpha)(1-\phi)}\log X_{it}+\frac{\alpha}{1-\alpha}\log\left(\frac{K_{it}}{Y_{it}}\right)+\frac{1}{1-\alpha}\gamma_{it}\log T-\frac{1}{1-\alpha}\gamma_{it}\log L_{it}+\varepsilon_{it} \tag{3.29}$$

where the subscript t on γ indicates that the output elasticity of land is not estimated from the data, but imposed using a measure of the land share. This equation implies that, out of steady state, if $\log y_{it}$, $\log X_{it}$, $\log\left(\frac{K_{it}}{Y_{it}}\right)$, and $\log L_{it}$ exhibit $I(1)$ behavior, and if γ_{it} is also non-stationary, then (for semi-endogenous growth theory to hold) $\log y_{it}$, $\log X_{it}$, $\log\left(\frac{K_{it}}{Y_{it}}\right)$, γ_{it} and $\gamma_{it}\log L_{it}$ must be cointegrated with a (single) cointegrating vector $(1, -\frac{\lambda}{(1-\alpha)(1-\phi)}, -\frac{\alpha}{1-\alpha}, -\frac{1}{1-\alpha}, \frac{1}{1-\alpha})$.

3.4.2. Findings

All studies that are based on one or more of the equations discussed above reject semi-endogenous growth in favor of Schumpeterian growth.

Minniti and Venturini (2017) estimate a distributed lag version of equation (3.22) and find no significant association between employment growth and output (value-added) per worker growth, whereas the share of R&D workers in employment is significantly and positively related to the growth rate of output per worker.

Laincz and Peretto (2006) use data on the growth rates of output per capita and private business productivity (as measures of g_y), employment (as a measure of L), R&D personnel (as a measure of X), and the number of establishments (as a measure of Q) and find that the ratio of L_t to Q_t is stationary. In addition, their results suggest that, although the unit root null hypothesis cannot be rejected for the X_t/Q_t ratio by the ADF test, cointegration exists between X_t and Q_t and between L_t and Q_t. Finally, using distributed lag versions of equations (3.23)-(3.26), they find that the b_1 coefficients are insignificant (with one exception), whereas the b_2 coefficients are significant and have a positive sign.

Madsen et al. (2010b) provide evidence of stationarity of the logarithmic ratio of X_t to Q_t. Consistent with this evidence in favor of Schumpeterian growth models, they also find that $\log X_t$ and $\log Q_t$ are cointegrated (although the estimated cointegrating vectors differ from the theoretically suggested value of $(1, -1)$). In addition, they find up to four cointegrating vectors among the variables in equation (3.29), depending on the period (and the test) chosen, and interpret this as "inconsistent with the prediction of only one cointegrated relationship among the variables by semi-endogenous growth theory" (p. 277). Moreover, they estimate equation (3.29) for the entire period and for sub-periods and find that the estimates differ from the theoretical values suggested in equation (3.29). Finally, they report estimates of (a lagged version of) equation (3.28) that show that the coefficients on (lagged) g_{xt} are always insignificant and sometimes have the "wrong" (negative) sign, whereas the coefficients on (lagged) $\log\frac{X_t}{Q_t}$ always have the "correct" (positive) sign and are mostly significant.

3.4.3. Critique

As discussed above, a necessary precondition for the validity of Schumpeterian growth theory is the stationarity of (the log of) R&D intensity, given that per capita output or labor productivity growth is generally found to be stationary. However, Laincz and Peretto (2006) find some evidence that the X_t/Q_t ratio has a unit root;[21] Madsen et al. (2010b) find, on the one hand, that the logarithmic ratio of X_t to Q_t is stationary and that $\log X_t$ is cointegrated with $\log Q_t$, while, on the other hand, they present estimates of the cointegrating relationship between $\log X_t$ and $\log Q_t$ showing that there is no one-to-one relationship between these variables (implying that the logarithmic ratio of X_t to Q_t is non-stationary); and Minniti and Venturini (2017) report (in their online appendix) mixed evidence of stationarity for the growth rate of labor productivity and strong evidence that R&D intensity is non-stationary.

If, however, R&D intensity is non-stationary but per capita (or per worker) growth is stationary, then equations (3.22), (3.24), (3.27), and (3.28) represent unbalanced regressions. As discussed in Section 3.3.3, the true coefficient on the non-stationary variable is necessarily zero in an unbalanced regression, implying that in studies that estimate unbalanced regressions and report statistically significant coefficients, the significance is necessarily spurious. Thus, it cannot completely be ruled out that at least some of the correlation between the growth rate of output per capita/worker and (the log of) $\frac{X_{it}}{Q_{it}}$ found in the studies above is spurious.

Similar caution should be exercised regarding the evidence against semi-endogenous growth models. The reason is that, as discussed in Section 3.3.3, models specified in growth rates of the variables—such as equations (3.22), (3.23), (3.25), and (3.28)—are misspecified if (as implied by equations (2.24), (2.25), (2.26), and (3.29)) there is a long-run cointegrating relationship between the (log-)levels of the variables. While Minniti and Venturini (2017) find, somewhat surprisingly, that $g_{y,it}$, $g_{L,it}$, and $\frac{X_{it}}{Q_{it}}$ are cointegrated (which would imply that their model is not misspecified due to the omission of the error-correction term),[22] Laincz and Peretto (2006), however, do not tests whether the log of GDP per capita/worker and the log of research effort or the log of employment are cointegrated, implying that their results may be misleading because they do not include the

[21] Laincz and Peretto (2006, p. 280) conclude that "the evidence is quite strong that average establishment size and R&D workers per establishment are stationary, trendless variables […]." However, the ADF test fails to reject the unit root null for the number of R&D workers per establishment.

[22] Minniti and Venturini (2017) find, using a small number of lags, that the growth rate of labor productivity is stationary. However, using a larger number of lags, they find that $g_{y,it}$ has a unit root. They also find that $g_{L,it}$ and $\frac{X_{it}}{Q_{it}}$ have a unit root, regardless of the number of lags included. Based on these findings, the authors treat $g_{y,it}$, and $g_{L,it}$, and $\frac{X_{it}}{Q_{it}}$ as $I(1)$ and test these variables for cointegration.

error-correction term capturing the potential cointegrating relationship between $\log y_{it}$ and $\log X_{it}$ or $\log L_{it}$.

Madsen et al. (2010b), in contrast, test for cointegration among the variables in equation (3.29), and find up to four cointegrating vectors, but they make no attempt to identify the cointegrating vectors; rather, they impose the restriction of one cointegrating vector over all estimation periods (although the evidence for one cointegrating relationship is weak), and justify this restriction by claiming that semi-endogenous growth theory predicts only one cointegrating vector between $\log y_t$, $\log X_t$, $\log\left(\frac{K_t}{Y_t}\right)$, γ_t, and $\gamma_t \log L_t$. However, while this claim holds for situations out of steady state, semi-endogenous theory implies (under the assumption of non-stationarity of γ_t) that in steady state, where the ratio of K_t to Y_t is constant, there are two cointegrating relationships among the five variables: one cointegrating relationship representing the stationary variable $\log\left(\frac{K_t}{Y_t}\right)$, and the other cointegrating relationship involving the other variables. In case that γ_t is stationary, it is even possible to find four cointegrating vectors in steady state: one cointegrating vector may simply reflect the stationarity of γ_t; $\log\left(\frac{K_t}{Y_t}\right)$ may account for the second stationary relationship; a third cointegrating relationship may exist between $\log X_t$ and $\gamma_t \log L_t$; and the fourth cointegrating relationship may involve $\log y_t$ and $\log X_t$. Thus, it can be argued that the evidence provided in Madsen et al. (2010b) is not sufficient to reject semi-endogenous growth with confidence.

These considerations point to a general problem with attempts to test semi-endogenous growth models using the Type III approach: The correct specification depends on whether the economy is in or out of steady state. To further illustrate this problem, recall that semi-endogenous growth models imply, according to equation (2.24), that we can write output per worker along a *constant* growth path as a function of research effort, the capital-output ratio, and human capital per worker. However, along a *balanced* growth path, output per worker is determined only by research effort, as equation (2.25) shows. Within the Type III approach, there are thus two possible empirical specifications for models of semi-endogenous growth whose correctness depends on whether or not economic growth is balanced. In contrast, the correctness of the specifications of the Type II approach does not depend on whether the economy is on a balanced or on a constant growth path. Similarly, the specifications of the Type I approach are correct regardless of whether the economy is in or out of steady state (but they all potentially suffer from unbalanced regression problems, as discussed in Section 3.2.3).

Finally, because λ is captured as part of the estimated coefficients, the Type III approach can be criticized for its failure to estimate the magnitude of duplication externalities.

3.5. The study by Sedgley and Elmslie (2010)

One study that does not fit into the above categories is that of Sedgley and Elmslie (2010). The authors provide a theoretical framework in which, assuming a constant R&D intensity, the steady-state growth rate of output per worker is proportional to the rate of population growth only in semi-endogenous growth models, while the capital-labor ratio and knowledge grow at the same rate along a balanced growth path in all R&D-based growth models (including first-generation R&D models). Hence, their empirically testable implications are as follows: If knowledge growth, capital per capita/worker growth, and population or employment growth are $I(1)$ processes, then semi-endogenous growth theory predicts two cointegrating relations among the three variables, whereas fully endogenous growth models (including Schumpeterian and first-generation R&D models) predict that only knowledge growth and capital per capita/worker growth are cointegrated.

Using the growth rate of the stock of patents as a measure of the growth rate of the knowledge stock, they find that the three variables exhibit $I(1)$ behavior. They also find cointegration between the growth rates of knowledge and capital per capita/worker, whereas they find no evidence of two cointegrating vectors, and interpret these findings as evidence in favor of fully endogenous growth models and against semi-endogenous growth models.

One problem with the study by Sedgley and Elmslie (2010), however, is that it does not explicitly test the validity of Schumpeterian theory. It assumes, rather than tests, the constancy or stationarity of R&D intensity. If R&D intensity is constant, then one would expect from Schumpeterian theory that the growth rate of the stock of knowledge is stationary. However, they find that their measure of the growth rate of the stock of knowledge, the growth rate of the stock of patents, is non-stationary, which inconsistent with their assumption of the stationarity of R&D intensity if Schumpeterian theory is valid.

Their finding of a unit root in the growth rate of the patent stock is consistent with Schumpeterian growth if R&D intensity has a unit root; but if R&D intensity exhibits $I(1)$ behavior, then R&D intensity should be cointegrated with the growth rate of the patent stock for Schumpeterian theory to hold.[23] However, they test neither for a unit root in R&D intensity nor for cointegration between R&D intensity and growth of the patent stock. Thus, it can be argued that their findings can be interpreted as evidence against semi-endogenous growth theory, but they do not provide any evidence in favor of Schumpeterian growth theory (or first-generation theory);

[23] Their finding of a unit root in the growth rate of the patent stock and their assumption of the stationary R&D intensity may be consistent with first-generation R&D models. However, since the steady-state growth rate of knowledge depends on population size and R&D intensity in these models (see, e.g., Jones, 1999), knowledge growth should be cointegrated with population size for these models to hold. Sedgley and Elmslie (2010) find that the growth rate of the population is an $I(1)$ process, implying that population size is $I(2)$. Given their finding that the growth rate of the patent stock is $I(1)$, it is not possible that the growth rate of the patent stock cointegrates with population size. Thus, their results are inconsistent with the first generation of R&D-based models if R&D intensity is stationary.

more specifically, their finding of cointegration between the growth rate of the patent stock and the growth rate of the capital-labor ratio is also consistent with exogenous growth models, and thus not sufficient to accept fully endogenous growth theory, because capital per worker grows at the rate of technical progress along a balanced growth path also in exogenous growth models.

Another, more fundamental problem associated with Sedgley and Elmslie's (2010) analysis is the use of the growth rate of the patent stock as a measure of the growth rate of the knowledge stock. As discussed in Section 3.2.3., the patent stock is probably a poor proxy for the stock of knowledge for two related reasons: the first is that the growth rate of output per worker in the United States (at least after 1980) is much smaller than one would expect on the basis of patent stock data if the patent stock were a good proxy for the knowledge stock; the second is that the growth rate of the patent stock exhibits $I(1)$ behavior (as found in this study and by Sedgley and Elmslie (2010)), while the growth rate of GDP per worker is typically found to be stationary; however, all R&D-based growth models suggest that both output per worker growth and knowledge growth should be stationary (or non-stationary). In this sense, their finding of a unit root in the growth rate of the patent stock can be interpreted either as inconsistent with the predictions of all R&D-based growth models or as an indication that the patent stock is not a good proxy for the stock of knowledge.[24]

3.6. Summary

In summary, it can be said that of the 17 studies reviewed, nine provide evidence in favor of Schumpeterian growth models; three support the hybrid semi-endogenous model; three present results that are largely inconsistent with Schumpeterian growth, semi-endogenous growth, and hybrid semi-endogenous growth; one provides results that can be interpreted as evidence against semi-endogenous growth models (but not necessarily as evidence of Schumpeterian growth); and one supports semi-endogenous growth.

However, all of these studies suffer from at least one of six major limitations:

(1) The majority of the studies use patent data to measure the flow of knowledge and/or to construct measures of the level and/or the growth rate of the stock of knowledge. A comparison of the evolution of the growth rate of GDP per capita with the evolution of the growth rate of the patent stock, however, suggests that the number/stock of patents is probably a poor measure of the flow/stock of knowledge (at least from the 1980s).

[24] Sedgley and Elmslie (2010) admit themselves that growth in per capita/worker output cannot be caused by growth in the stock of patents since the former is typically found to be stationary (while the latter is $I(1)$). They argue that it is necessary to use a linear combination of at least two $I(1)$ series to explain the stationary behavior of per capita output growth and suggest that the other $I(1)$ variable could be the growth of the capital stock. However, we are not aware of any R&D-based growth model that predicts a cointegrating relationship between the growth rates of capital and knowledge.

(2) Many of the studies pay little to no attention to the possibility of spurious regressions due to non-stationary data for the dependent and/or the independent variables. Hence, some of the reported significance levels may be misleading, and the evidence from these studies for Schumpeterian or (hybrid) semi-endogenous growth may be false.

(3) Some of the studies use time series on the number of researchers and/or R&D expenditures with large, interpolated data gaps. If there are significant discrepancies between the interpolated and true values of these variables, cointegration tests can be biased in favor of the null hypothesis of no cointegration between the log of the stock of knowledge and the log of research effort and thus toward rejecting semi-endogenous growth.

(4) The studies often use simple growth regressions to test semi-endogenous growth models. If there is a long-run (cointegrating) relationship between the log-levels of the variables, such regressions, however, are misspecified by the omission of the error-correction term that captures the long-run relationship.

(5) Most of the studies employ R&D expenditures as a measure of research effort. A potential problem is that, since there are no long time series on the price of R&D, these studies use an overall deflator instead of an R&D-specific deflator to convert nominal R&D expenditures to real R&D expenditures. Thus, failure to find empirical support for semi-endogenous growth could also (in part) be due to measurement error in the data on real R&D expenditures.

(6) Several studies use unit root tests to examine the stationarity of R&D intensity. These studies may suffer from the problem that standard unit root tests may falsely reject the unit root null hypothesis when the lag length is incorrectly specified to be too small. Thus, these studies may incorrectly conclude from their unit root test that Schumpeterian theory is supported by the data.

Finally, a common limitation of studies using the Type I approach is that the underlying regressions are all likely to be unbalanced in the sense that the order of integration of the dependent knowledge flow variable is smaller than the order of integration of the explanatory knowledge stock variable, which can lead to biased results; general limitations of the Type III approach are that the correct specification to test semi-endogenous growth theory depends on whether the economy is in or out of steady state, which can lead to uncertainty in the results, and that the approach fails to provide estimates of λ (estimates that vary widely across studies using Type I and II).

The following analysis aims to overcome these limitations by (*i*) applying unit root and cointegration methods within the Type II approach, (*ii*) using proper lag selection procedures, (*iii*) employing TFP as a measure of the stock of knowledge and the number of researchers as a measure of research effort, and (*iv*) using non-interpolated data on the number of researchers. Unit root and cointegration methods not only provide tests of the long-run predictions of (hybrid) semi-endogenous and Schumpeterian growth models, but also help to avoid potential problems associated with spurious regressions, unbalanced regressions, and simple growth regressions.

To increase the comparability of our results with those in the studies reviewed here, we conduct both a time-series analysis for the United States and a panel data analysis of OECD countries (excluding the United States), as already noted in the Introduction.

4. New empirical evidence on the validity of semi-endogenous and Schumpeterian growth models

4.1. Data

The stock of knowledge is measured by TFP. Following common practice, we calculate TFP as the residual from the production function (2.1), using employment times average hours times human capital per worker as the measure of human capital-augmented labor input and assuming $\alpha = 1/3$. All data used to calculate TFP are from the Penn World Tables (PWT) version 9.0 (available at https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt9.0).[25]

Our measure of research input is the number of full-time equivalent researchers, L_{Ait}; total employment, measured by the number of employed persons, L_{it}, is used as a measure of scale; and R&D intensity is measured by the share of researchers in total employment (researchers per 1000 employees), L_{Ait}/L_{it}. Data on these variables are from the Main Science and Technology Indicators (MSTI) of the OECD (available at https://stats.oecd.org/Index.aspx?DataSetCode=MSTI_PUB#).

Since the PWT data end in 2014 and the MSTI data start in 1981, our sample covers the period from 1981 to 2014. For the United States, complete data are available for this sample period, implying that our time-series analysis is based on 34 observations. For our panel analysis, we include all countries outside the United States with complete time series and at least 20 time-series observations, resulting in an unbalanced panel of 500 observations from 19 OECD countries.

The growth rate of TFP and the log of the share of researchers in total employment between 1981 and 2014 for the 20 countries in our study are plotted in Figure 1. As already noted in Section 3.3.3, the TFP growth rates in all countries show no persistent increase during this period, whereas the log of the number of researchers as a proportion of employment shows an upward trend for each

[25] The Penn World Tables 9.0 contains its own measure of TFP, which is based on a translog production function in which the labor share varies across countries and across time. However, as argued by Jones (2016), such a measure is problematic because it implies that countries and years with the same inputs and the same level of TFP will have different outputs. In fact, it is still debated whether the labor share is approximately constant across time and space (with a value of about 2/3). While Karabarbounis and Neiman (2014) document a secular decline in the labor share in most advanced countries since the early 1980s, Cette et al. (2020) challenge this finding and demonstrate that, when corrected for measurement error, the labor share of advanced economies does not follow a secular trend. Therefore, we follow the common practice of assuming $\alpha = 1/3$.

country. This visual inspection of the data already suggests that there is no evidence to support Schumpeterian theory. In the next sections, we investigate this issue further.

4.2. Unit root tests

To examine the time-series properties of $\log A_t$, $\log L_{At}$, $\log L_t$, and $\log(L_{At}/L_t)$ for the United States, we use the augmented Dickey-Fuller (*ADF*) test, the *DF-GLS* test of Elliott et al. (1996), and the MZ_{α}^{GL}, MZ_t^{GLS}, MSB^{GLS}, and MP_T^{GLS} tests of Ng and Perron (2001). The lag length for the ADF test is chosen by the general-to-specific criterion of Hall (1994). For all other tests, we employ the modified Akaike information criterion (MAIC) of Ng and Perron (2001). Following the recommendation of Perron and Qu (2006), we calculate the MAIC using OLS (rather than GLS) detrended data. The results based on these lag selection procedures are presented in Table 2. All six tests fail to reject the unit root null for the levels of each of the variables. In contrast, the null hypothesis of a unit root for the first differences is rejected, indicating that we have stationarity in first differences and each of the four variables can be regarded as $I(1)$. Thus, the results of the unit root tests for the United States provide evidence against Schumpeterian growth models.

To test for unit roots in our panel data, we use a second-generation panel unit root tests: the cross-sectionally augmented panel unit root test (CIPS) of Pesaran (2007). Panel unit root tests of the first-generation can lead to spurious results if a significant degree of residual cross-sectional dependence exist (but is ignored).[26] In fact, the *p*-values of the cross-sectional dependence (CD) test of Pesaran (2004), reported in Table 3, indicate that both the levels and first differences of the series are not cross-sectionally independent, implying that error cross-sectional dependence can be a serious problem (if not controlled). The CIPS test, which is based on averaging individual ADF statistics, is designed to control for cross-sectional dependence in the residuals by augmenting the ADF regressions with cross-sectional averages. Table 2 presents the *p*-values of the CIPS tests for different lags $k = 0, \ldots, 3$. Regardless of the lag length chosen, the null hypothesis of a unit root cannot be rejected for all series in levels, but can be rejected for all in their first differences. Thus, our panel unit root tests, like our unit root tests for the United States, reject Schumpeterian growth models.

4.3. Cointegration tests

We next test for cointegration between $\log L_{At}$ and $\log L_t$ and between $\log A_t$ and $\log L_{At}$ for the United States using the standard Engle and Granger (1987), Phillips and Ouliaris (1990), and Pesaran et al. (2001) methods. The results of these tests, which are reported in Panels A and B of Table 4, show

[26] Cross-sectional dependence may be due to common factors that affect all countries (but not necessarily with the same magnitude) and/or spatial spillover effects across subsets of countries.

that the null hypothesis of no cointegration between $\log L_{At}$ and $\log L_t$ cannot be rejected, while the null hypothesis of no cointegration between $\log A_t$ and $\log L_{At}$ can be rejected. Thus, we again find evidence against Schumpeterian models. In contrast, the data support the implication of semi-endogenous growth models that if $\log A_t$ and $\log L_{At}$ are $I(1)$, these variables must be cointegrated. In addition, the finding that $\log L_{At}$ is cointegrated with $\log A_t$ but not cointegrated with $\log L_t$ implies that the hybrid semi-endogenous is inconsistent with the data (as discussed in Section 3.3.1).

To test for cointegration in our panel, we use the standard panel and group ADF and PP test statistics of Pedroni (1999). Given that these tests, which assume cross-sectionally independent residuals, suffer from size distortions in the presence of error cross-sectional dependence, we demean the data by subtracting the average value of $x_t = \left(\sum_{i=1}^{N} x_{it}\right)/N$ from each x_{it} in each period t, $x_{it} - \left(\sum_{i=1}^{N} x_{it}\right)/N$, and use the demeaned data in place of the original data. In addition, we use the error-correction model (ECM) t-test of Gengenbach et al. (2016), which, like the CIPS test, accounts for cross-sectional dependence via the use of cross-sectional averages (and is therefore applied to the raw data).

The results of our panel cointegration tests are given in Panels A and B of Table 5. In Panel A, we see that the tests fail to reject the null of no cointegration between $\log L_{Ait}$ and $\log L_{it}$. In contrast, in Panel B, we see that the null of no cointegration between $\log A_{it}$ and $\log L_{Ait}$ is rejected by all tests. Thus, like our time-series results for the United States, our panel results provide evidence consistent with semi-endogenous growth.

4.4. Cointegrating relationship

To estimate the cointegrating parameter $\frac{\lambda}{1-\phi}$ for the United States, we use the dynamic OLS (DOLS) estimator of Stock and Watson (1993) and the fully modified ordinary least squares (FMOLS) estimator of Phillips and Hansen (1990). The estimates are reported in Table 6. In the Table, we also report Hansen's (1992) L_c statistic for parameter instability, which is also a test of the null of cointegration. Since the L_c statistics are not significant at the 20%, both the null of parameter stability and the null of cointegration cannot be rejected, both for the DOLS regression and the FMOLS regression. Thus, we again find evidence for the United States that $\log A_t$ and $\log L_{At}$ are cointegrated. In addition, we find a positive and highly significant coefficient on $\log L_{At}$ in both regressions for the United States, as semi-endogenous growth models predict.

To provide panel estimates of $\frac{\lambda}{1-\phi}$, we use the panel DOLS (PDOLS) and panel FMOLS (PFMOLS) estimators of Kao and Chiang (2000). Both estimators are applied to the demeaned data to account for error cross-sectional dependence. The estimation results are presented in Table 7, along with the results of the CD test applied to the residuals from the PDOLS and PFMOLS regressions. As can be seen, the CD test indicates that the PDOLS and PFMOLS estimates do not suffer from error cross-sectional dependence, and the estimated coefficients on $\log L_{Ait}$ are highly significant. Consistent with the results for the United States, our panel estimates thus provide evidence supporting semi-endogenous growth models, although the panel estimates of $\frac{\lambda}{1-\phi}$ are smaller than the time-series ones.

4.5. Error-correction models

Having identified the long-run relationships between $\log A_{it}$ and $\log L_{Ait}$ for the United States and our panel of 19 countries, we are able to construct error-correction models that allow us to estimate λ and ϕ.

Since the DOLS estimator performs better in small samples than the FMOLS estimator (see, e.g., Stock and Watson, 1993), we use the DOLS estimates of the cointegrating relationships to calculate the error-correction terms as $ec_{1it} = \log A_{it} - \widehat{\frac{\lambda}{1-\phi}}\log L_{Ait}$ and include these terms (lagged one period) in regression models of the form (3.21). To account for the possibility that $\Delta\log A_{it}$ responds to $\Delta\log L_{Ait}$ with a lag, up to four lagged values of these variables are also included.

However, while estimates of cointegrating relationships are robust to omitted variables, the problem with this specification is that it may produce biased estimates of λ if unobserved business-cycle effects increase both the growth rate of TFP and the growth rate of research effort. More specifically, the estimated coefficient on the current value of $\Delta\log L_{Ait}$ may be biased upward because of the well-known procyclicality of both TFP and R&D. When lagged values of $\Delta\log L_{Ait}$ are included in equation (3.21), however, the coefficient on lagged $\Delta\log L_{Ait}$ may be biased downward because the current business cycle may be negatively correlated with lagged growth in R&D activities but positively correlated with current growth in TFP. This source of bias is less of a concern for the estimation of the coefficient on ec_{1it-1} because it measures the single-period response of the dependent variable to departures from equilibrium and must therefore be negative and significant (if $\log A_{it}$ and $\log L_{Ait}$ are cointegrated and $\log A_{it}$ is not weakly exogenous). In addition, the error-correction term by construction should be weakly correlated or uncorrelated with business-cycle effects that stimulate both productivity growth and research. Therefore, we use the estimated coefficients on ec_{1it-1} to derive estimates of ϕ (using the delta method), but we do not use the short-run dynamics of the models to identify λ.

To identify λ, we make use of the facts that parameter estimates are robust to endogeneity assumptions when variables are cointegrated and that equation (3.21) can be rewritten as

$$\Delta \log A_{it} = c_{1i} + \lambda \left[\log X_{it\text{-}1} - \frac{1}{\frac{\lambda}{1-\phi}} \log A_{it\text{-}1} - \frac{1}{\frac{\lambda}{1-\phi}} c_i \right] + \lambda \Delta \log X_{it} + \varepsilon_{it} \tag{3.30}$$

This allows us to alternatively calculate the error-correction terms as $ec_{2it} = \log L_{Ait} - (1/\widehat{\frac{\lambda}{1-\phi}}) \log A_{it}$ using the reciprocals of the DOLS estimates of $\frac{\lambda}{1-\phi}$, and to use these terms alternatively in the regressions of $\Delta\log A_{it}$ on (country-specific) constant(s), up to four lagged values of $\Delta\log A_{it}$, and both current and up to four lagged values of $\Delta\log L_{Ait}$, In doing so, we can estimate the values for λ through the coefficients on $ec_{2it\text{-}1}$.

Table 8 reports the error-correction model results for the United States using the general-to-specific approach. Both error-correction terms have the expected signs and are highly statistically significant, which is consistent with the finding that $\log A_t$ and $\log L_{At}$ are cointegrated. In addition, the coefficient on $ec_{1t\text{-}1}$ implies a scale parameter of 0.656, while the coefficient on $ec_{2t\text{-}1}$ yields a value for the duplication parameter of 0.129. This value differs markedly from the λ estimate of 0.055 implied by the coefficient on $\Delta\log L_{At\text{-}2}$ (which is significant only 10% level), suggesting that the first-difference estimate of λ is biased by business cycle effects, as discussed above.

Table 9 presents our panel results (also using the general-to-specific approach). As in Table 8, the error correction terms are highly statistically significant. However, the estimate of ϕ is larger than its counterpart in Table 8, whereas the estimated value of λ is smaller than its counterpart in Table 8. More specifically, the estimated value of ϕ implied by the coefficient on ec_{1it-1} is 0.930, and the coefficient on ec_{2it-1} implies a value of λ of 0.012, which again differs from the value of the coefficient on the first difference of $\log L_{Ait}$.

All together, these results suggest that semi-endogenous theory is empirically valid, but there are strong negative duplication externalities resulting from competition for new ideas. These externalities appear to be stronger in countries outside the United States.

5. Conclusions

Our review and analysis lead to two main conclusions: First, there are two measurement issues that plague most studies. The first is that the evolution of per capita output implies an evolution of the stock of knowledge that is inconsistent with the evolution of the number/stock of patents (at least since the 1980s). It is therefore doubtful whether studies based on patent data can properly test the different assumptions of semi-endogenous and Schumpeterian growth models. A second measurement issue relates to the use of an overall price deflator to deflate R&D expenditures, which may lead to measurement error in real R&D expenditures. In fact, we find (not reported) that neither semi-endogenous growth theory nor Schumpeterian growth theory is supported by our analysis when real R&D expenditures based on the GDP deflator (from the MSTI) are used as a measure of R&D activity. We therefore recommend the use of TFP as a measure of the stock of knowledge and the number of researchers as a measure of research effort.

Second, on the one hand, there are possible reasons why studies might fail to find evidence of semi-endogenous growth, including bias resulting from estimating difference models with cointegrated variables without an error-correction term and/or mismeasurement of R&D inputs due to interpolation and deflation errors. On the one hand, there are possible reasons why studies might find spurious evidence for Schumpeterian growth, including spurious rejections of the unit root hypothesis for R&D intensity due to incorrectly specified lags and/or spurious regression problems associated with non-stationary data. In contrast to most previous studies, we find strong evidence in favor semi-endogenous growth, suggesting indeed that at least some of the evidence against semi-endogenous growth is spurious.

References

Abdih, Y., Joutz, F., 2006. Relating the knowledge production function to total factor productivity: an endogenous growth puzzle. *IMF Staff Papers*, 53, 242–27.

Ang, J.B., 2010a. Financial reforms, patent protection and knowledge accumulation in India. *World Development*, 38, 1070-1081.

Ang, J.B., 2010b. Research, technological change and financial liberalization in South Korea. *Journal of Macroeconomics*, 32, 457-468.

Ang, J.B., 2014. Innovation and financial liberalization. *Journal of Banking and Finance*, 47, 214-229.

Ang, J.B., Madsen, J.B., 2011. Can second-generation endogenous growth models explain the productivity trends and knowledge production in the Asian miracle economies? *Review of Economics and Statistics*, 93, 1360-1373.

Ang, J.B., Madsen, J.B., 2015. What drives ideas production across the world? *Macroeconomic Dynamics*, 19, 2015, 79-115.

Ang, J.B., Madsen, J.B., 2016. Finance-led growth in the OECD since the nineteenth century: How does financial development transmit to growth? *Review of Economics and Statistics*, 98, 552-572.

Barcenilla-Visús, S., López-Pueyo, C., Sanaú-Villarroya, J., 2014. Semi-endogenous versus fully endogenous growth theory: A sectoral approach. *Journal of Applied Economics*, 17, 1-30.

Barlevy, G., 2007. On the cyclicality of research and development. *American Economic Review*, 97, 1131-1164.

Bekiros, S., Sjö, B., Sweeney, R.J., 2018. Pitfalls in cross-section studies with integrated regressors: A survey and new developments. *Journal of Economic Surveys*, 32, 1045-1073.

Bond-Smith, S., 2019. The decades-long dispute over scale effects in the theory of economic growth. *Journal of Economic Surveys*, 33, 1-29.

Cette, G., Koehl, L., Philippon, T., 2020. Labor share. *Economics Letters*, 188, 108979

Elliott, G., Rothenberg, T.J., and James Stock, J., 1996. Efficient tests for an autoregressive unit root. *Econometrica*, 64, 813-836.

Engle, R.E., Granger, C.W.J., 1987. Cointegration and error-correction: representation, estimation, and testing. *Econometrica*, 55, 251-276.

Fedderke, J.W., Liu, Y., 2017. Schumpeterian and semi-endogenous productivity growth explanations. *Economics of Transition*, 25, 111-137.

Furman, J.L., Porter, M.E., Stern, S., 2002. The determinants of national innovative capacity. *Research Policy*, 31, 899-933.

Gengenbach, C., Urbain, J.-P., Westerlund, J., 2016. Error correction testing in panels with common stochastic trends. *Journal of Applied Econometrics*, 31, 982-1004.

Ha, J., Howitt, P., 2007. Accounting for trends in productivity and R&D: A Schumpeterian critique of semi-endogenous growth theory. *Journal of Money, Credit and Banking*, 39, 733-77.

Haldrup, N., 1998. An econometric analysis of I(2) variables. *Journal of Economic Surveys*, 12, 595-650.

Hall, A., 1994. Testing for a unit root in time series with pretest data-based model selection. *Journal of Business and Economic Statistics*, 12, 461-470.

Hansen, B.E., 1992. Tests for parameter instability in regressions with 1(1) processes. *Journal of Business and Economic Statistics*, 10, 321-335.

Howitt, P., 1999. Steady endogenous growth with population and R&D inputs growing. *Journal of Political Economy*, 107, 715-730.

Hu, M.-C., Mathews, J.A., 2005. National innovative capacity in East Asia. *Research Policy*, 34, 1322-1349.

Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231-254.

Jones, C.I., 1995. R&D-based models of economic growth. *Journal of Political Economy*, 103, 759-784.

Jones, C.I., 1999. Growth: with or without scale effects? *American Economic Review*, 89, 139-144.

Jones, C.I., 2002. Sources of U.S. economic growth in a world of ideas. *American Economic Review*, 92, 220-39.

Jones, C.I., 2016. The facts of economic growth. In: Taylor, J.B., Uhlig, H. (eds.), Handbook of Macroeconomics, Saint Louis: Elsevier Science & Technology, 3-69.

Kao, C., 1999. Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, 90, 1-44.

Kao, C., Chiang, M., 2000. On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics*, 15, 179-222.

Karabarbounis, L., Neiman, B., 2014. The global decline of the labor share. *Quarterly Journal of Economics*, 129, 61-103.

Kortum, S., 1997. Research, patenting, and technological change. *Econometrica*, 65, 1389–1419.

Laincz, C.A., Peretto, P.F., 2006. Scale effects in endogenous growth theory: an error of aggregation not specification. *Journal of Economic Growth*, 11, 263-288.

Luintel, K.B., Khan, M., 2009. Heterogeneous ideas production and endogenous growth: an empirical investigation. *Canadian Journal of Economics*, 42, S.1176-1205.

Luintel, K.B., Khan, M., 2017. Ideas production and international knowledge spillovers: Digging deeper into emerging countries. *Research Policy*, 46, 1738-1754.

Madsen, J., 2008. Semi-endogenous versus Schumpeterian growth models: testing the knowledge production function using international data. *Journal of Economic Growth*, 13, 1-26.

Madsen, J.B., Saxena, S., Ang, J.B., 2010a. The Indian growth miracle and endogenous growth. *Journal of Development Economics*, 93, 37-48.

Madsen, J.B., Ang, J.B., Banerjee, R., 2010b. Four centuries of British economic growth: the roles of technology and population. *Journal of Economic Growth*, 15, 263-290.

Minniti, A., Venturini, F., 2017. The long-run growth effects of R&D policy. *Research Policy*, 46, 316-326.

Neves, P.C., Sequeira, T., 2018. Spillovers in the production of knowledge: A meta-regression analysis. *Research Policy*, 47, 750-767.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69, 1519-1554.

Pedroni, P., 1999. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61, 653-670.

Peretto, P., 1998. Technological change and population growth. *Journal of Economic Growth*, 3, 283-311.

Perron, P., Qu, Z., 2006. A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters*, 12-19.

Pesaran M.H., 2004. General diagnostic tests for cross section dependence in panels. CESifo Working Paper Series no. 1229, CESifo Group, Munich.

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22, 265-312.

Pesaran, M.H., Shin, Y., Smith, R.J., 2001. Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16, 289-326.

Phillips, P., Hansen, B., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57, 99-125.

Phillips, P., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58, 165-93.

Saunoris, J.W., Payne, J.E., 2011. An empirical note on R&D growth models with regional implications. *Journal of Regional Analysis and Policy*, 41, 16-21.

Sedgley, N., Elmslie, B., 2010. Reinterpreting the Jones critique: A time series approach to testing and understanding idea driven growth models with transitional dynamics. *Journal of Macroeconomics*, 32, 103-117.

Segerstrom, P.S., 1998. Endogenous growth without scale effects. *American Economic Review*, 88, 1290-1310.

Stewart, C., 2011. A note on spurious significance in regressions involving I(0) and I(1) variables. *Empirical Economics*, 41, 565-571.

Stock, J.H.,Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61, 783-820.

Ulku, H., 2007. R&D, innovation and output: Evidence from OECD and non-OECD countries. *Applied Economics*, 39, 291-307.

Venturini, F., 2011. Looking into the black box of Schumpeterian growth theories: An empirical assessment of R&D races. Quaderni del Dipartimento di Economia, Finanza e Statistica, Quaderno n. 94, Università di Perugia.

Venturini, F., 2012a. Product variety, product quality, and evidence of Schumpeterian endogenous growth: A note. *Economics Letters*, 117, 74-77.

Venturini, F., 2012b. Looking into the black box of Schumpeterian growth theories: An empirical assessment of R&D races. *European Economic Review*, 56, 1530-1545.

Young, A., 1998. Growth without scale effects. *Journal of Political Economy*, 106, 41-63.

Figure 1 Growth rate of TFP (—) and log of the share of researchers in total employment (–) between 1981 and 2014

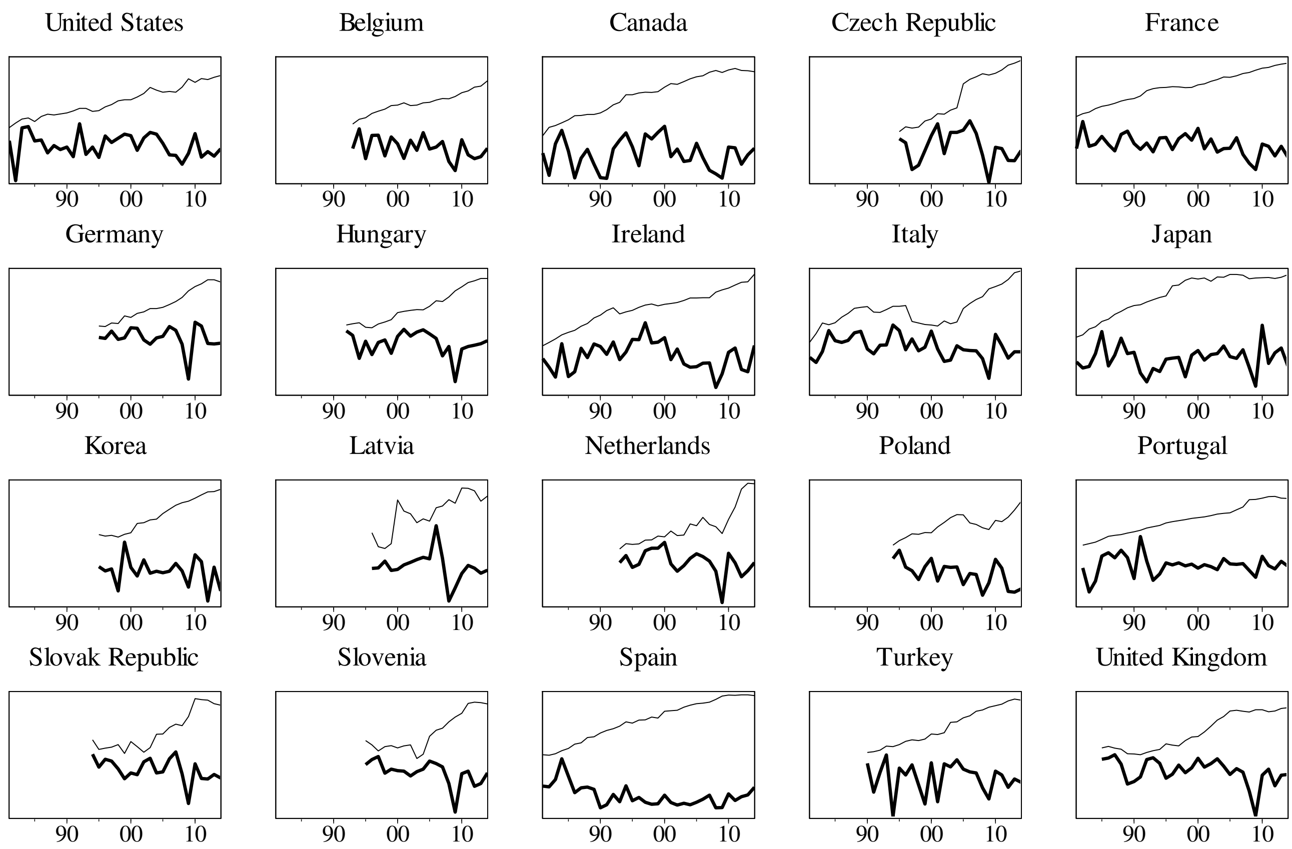

Table 1 Summary of studies testing semi-endogenous versus Schumpeterian growth models

(1) Study [Approach]	(2) Main variables	(3) Type of analysis and number of countries, industries or firms [Period]	(4) Econometric methods	(5) Conclusion
Furman et al. (2002) [Type I]	$\dot{A}$: *PatG*, *A*: *PatGStock*_0, *GDP*/*POP* *X*: L_A, *R&D* *Q*: *POP*	Panel data analysis covering 17 OECD countries [1973–1996]	FE, POLS	Results support the hybrid semi-endogenous model
Hu and Mathews (2005) [Type I]	$\dot{A}$: *PatG*, *A*: *PatGStock*_0, *GDP*/*POP* *X*: L_A, *R&D* *Q*: *POP*	Panel data analysis covering 5 Asian countries [1975–2000]	FE, POLS	Results are inconsistent with Schumpeterian growth and (hybrid) semi-endogenous growth
Laincz and Peretto (2006) [Type III]	g_y: growth rate of *GDP7POP*, growth rate of *PBP* g_L: growth rate of *L* g_X: growth rate of L_A *X*/*Q*: L_A /*NE* *L*/*Q*: *L*/*NE*	Time-series analysis for one country, the United States [1964–2001]	Unit root and cointegration tests, distributed lag regressions	Evidence in favor of Schumpeterian growth models
Ha and Howitt (2007) [Type II]	*A*: *TFP* *X*: L_A, L_AG5,[a] *R&D/TFP* *X*/*Q*: L_A/L, $L_A/(h\times L)$, *R&D*/(*TFP*×*L*), *R&D*/(*TFP*× *h*×*L*), *R&D*/*GDP*	Time-series analysis for one country, the United States [1950–2000 (1953–2000 for R&D)]	Unit root and cointegration tests, ML	Evidence in favor of Schumpeterian growth models
Madsen (2008) [Type II]	*A: TFP* *X: Pat, PatG, R&D* *Q: TStock_20, GDP, L* *X/Q: R&D/GDP, R&D/(TFP×L), R&D/(TFP× TStock_20), Pat/L, Pat/ TStock_20*	Panel data analysis covering 21 OECD countries [1965–2004] / 14 OECD countries [1900–2004] and time-series analysis for the United States [1922–2004], Japan [1953–2004], Australia [1940–2004], Germany [1870–2004], and Spain [1870–2004]	Panel and time series cointegration tests, DOLS, FE	Evidence in favor of Schumpeterian growth models
Luintel and Khan (2009) [Type I]	$\dot{A}$: *TPat* *A*: *PatGStock*_15 *X*: L_A *X*/*Q*: L_A/L	Panel data analysis covering 19 OECD countries [1981–2000]	GMM based on a dynamic heterogeneous panel model	Results support the hybrid semi-endogenous model
Madsen et al. (2010a) [Type II]	*A: TFP* *X*: L_A, *R&D, Pat, PatG* *X/Q*: L_A/L, $L_A/(h\times L)$, *R&D/(h×L), R&D/GDP, Pat/L, PatG/L*	Time-series analysis for one country, India [1950–2005], and panel data analysis covering 590 Indian firms [1993–2005]	Time series and panel unit root and cointegration tests, ML (aggregate data)[b], OLS with HAC standard errors, FE	Evidence in favor of Schumpeterian growth models

(1) Study [Approach]	(2) Main variables	(3) Type of analysis and number of countries, industries or firms [Period]	(4) Econometric methods	(5) Conclusion
Madsen et al. (2010b) [Type III]	g_y: growth rate of *GDP/POP* g_X: growth rate of *Pat* *X/Q*: *Pat/POP* *y*: *GDP/POP* *X*: *Pat*	Time-series analysis for one country, Britain (England and Wales) [1620–2006]	Unit root and cointegration tests, ML, OLS with HAC standard errors	Evidence in favor of Schumpeterian growth models
Sedgley and Elmslie (2010) [Unclassified]	g_A: growth rate of *PatGStock_0* g_{POP}: growth rate of *POP* g_L: growth rate of *L* g_k: growth rate of the capital stock per capita/worker	Time-series analysis for one country, the United States [1951–2000]	Unit root and cointegration tests, ML	Evidence against semi-endogenous growth models
Ang and Madsen (2011) [Type I and Type II]	$\dot{A}$: *Pat* *A*: *TFP*, *PatStock*_15 *X*: *R&D*, L_A *Q*: *GDP*, *L* *X/Q*: *R&D/GDP*, *R&D*/(*TFP*×*L*), *R&D*/(*PatStock*_15×*L*), L_A/L, $L_A/(h\times L)$	Panel data analysis covering 6 Asian countries [1953–2006]	Panel unit root and cointegration tests, SUR	Evidence in favor of Schumpeterian growth models
Saunoris and Payne (2011) [Type II]	*A*: *TFP* *X*: *R&D* *Q*: *GDP*, *TFP*×*h*×*L*, *TFP*×*L*	Time-series analysis for one country, the United States [1960–2007]	Unit root and cointegration tests, DOLS	Evidence in favor of Schumpeterian growth models
Venturini (2012a) [Type I]	$\dot{A}$: *Pat, FCPat, BCPat, PCPat* *A: PatStock_15* *X: R&DStock_15* *Q: SO*	Panel data analysis covering 20 US manufacturing industries [1975–2003]	Panel unit root and cointegration tests, DOLS	Results support the hybrid semi-endogenous model
Venturini (2012b) [Type II]	$\dot{A}_{it}/A_{it}$: *Pat*/*PatStock*_15 *A*: number of innovating firms (relative to *SO* or *L*) *X*: *R&D* (relative to *SO*), *R&DStock*_15 (relative to *SO*), L_A (relative to *L*) *X*/*Q*: *R&D*/*SO*, L_A/*L*, *R&D*/*L*, *R&DStock*_15/*L*	Panel data analysis covering 12 US manufacturing industries [1975–1996]	FE-IV	Evidence in favor of semi- endogenous growth models
Barcenilla-Visús et al. (2014) [Type II]	*A*: *TFP* *X*: *R&D*, *R&D/ TFP* *Q*: *L*, *SV*, *SH*, *TFP*×*SH* *X*/*Q*: *R&D*/(*TFP*×*SH*), *R&D*/*SV*	Panel data analysis covering 10 manufacturing industries in 6 OECD countries (Canada, Finland, France, Italy, Spain, and the United States) [1979–2001]	Panel unit root and cointegration tests, DOLS, FE, POLS	Results are inconsistent with Schumpeterian growth and (hybrid) semi-endogenous growth

(1) Study [Approach]	(2) Main variables	(3) Type of analysis and number of countries, industries or firms [Period]	(4) Econometric methods	(5) Conclusion
Ang and Madsen (2015) [Type I]	$\dot{A}$: *Pat*, *PatGF* *A*: *PatStock*_15, *PatStock*_5, *PatStock*_25 *X*: *R&D* *Q*: *GDP*	Panel data analysis covering 26 countries [1870–2010]	FGLS	Evidence in favor of Schumpeterian growth models
Minniti and Venturini (2017) [Type III]	g_y: growth rate of *SV/L* g_L: growth rate of *L* *X/Q*: L_A/L	Panel data analysis covering 20 US manufacturing industries [1975–2000]	Panel unit root and cointegration tests, CSDL	Evidence in favor of Schumpeterian growth models
Fedderke and Liu (2017) [Type I and Type II]	$\dot{A}$: change in *TFP* *A*: *TFP* *X*: *R&D/TFP* *Q*: *L*, *SH*, *TH*, *SV*, *GDP*, *Pat* *X/Q*: *R&D*/(*TFP*×*L*), *R&D*/(*TFP*×*SV*), *R&D*/(*TFP*×*GDP*), *R&D*/(*TFP*×*Pat*), *R&D*/(*TFP*×*TH*), *R&D*/(*TFP*×*SH*)	Panel data analysis covering 13 countries [1996–2010] / 25 manufacturing industries in South Africa [1973–1993] / 10 manufacturing sectors in 6 OECD countries (Canada, Finland, France, Italy, Spain, and the United States) [1979–2001] and time-series analysis for 25 manufacturing industries in South Africa [1973–1993] and 10 manufacturing sectors in 6 OECD countries [1979–2001]	Panel unit root tests (but no panel cointegration tests), PMG, GMM, time series unit root and cointegration tests, ML	Results are inconsistent with Schumpeterian growth and (hybrid) semi-endogenous growth

Notes: Abbreviations in column (2) are: *BCPat* = backward-citation-weighted patents, *FCPat* = forward-citation-weighted patents, *Firms* = number of innovating firms, *GDP* = real gross domestic product, *L* = total (sectoral) employment, L_A = number of researchers, L_AG5 = sum of the number of scientists and engineers engaged in R&D in the G-5 countries (France, West Germany, Japan, the United Kingdom, and the United States), *NE* = number of establishments, *Pat* = patent applications, *PatG* = patents granted, *PatGF* = patents granted to foreign residents, *PatStock_x* = stock of patent applications with a depreciation rate of x%, *PatGStock_x* = stock of granted patents with a depreciation rate of x%, *PBP* = private business productivity, *PCPat* = priority claims-weighted patents, *POP* = population size, *R&D* = real research and development expenditures, *R&DStock*_15 = stock of R&D expenditures with a depreciation rate of 15%, *SH* = sectoral hours worked, *SO* = real sectoral output, *SV* = real sectoral value added, *TH* = total working hours, *TPat* = triadic patents (granted by all three major patent offices, the USPTO, the European Patent Office, and the Japanese Patent Office), *TFP* = total factor productivity, *TStock*_20 = stock of trademarks with a depreciation rate of 20%.

Abbreviations in column (4) are: CSDL = cross-sectionally augmented distributed lag estimator, DOLS = dynamic ordinary least squares estimator, FE = fixed-effects estimator, FE-IV = fixed effects instrumental variables estimator, FGLS = feasible generalized least squares estimator, FMOLS = fully modified least squares estimator, GMM = generalized method of moments estimator, HAC = heteroskedasticity and autocorrelation consistent, ML = maximum likelihood estimator, POLS = pooled ordinary least squares estimator, PMG = pooled mean group estimator, SUR = seemingly unrelated estimator.

[a] Ha and Howitt (2007) use both the number of scientists and engineers engaged in R&D in the United States and the number of scientists and engineers engaged in R&D in the G-5 countries, L_AG5. The latter takes into account the possibility that much of the relevant input to U.S. productivity growth comes from ideas generated in other leading industrial countries.

[b] Madsen et al. (2010a) provide no information regarding the panel cointegration estimator they use for the firm level data

Table 2 Time-series unit root tests for the United States

	ADF	*DF–GLS*	MZ_{α}^{GLS}	MZ_{t}^{GLS}	MSB^{GLS}	MP_{T}^{GLS}
Levels						
$\log A_t$	-0.931	-0.187	0.667	0.520	0.780	42.095
$\log L_{At}$	-1.385	0.671	0.446	0.273	0.612	27.384
$\log L_t$	-2.613	-0.625	-0.691	-0.353	0.511	17.124
$\text{Log}(L_{At}/L_t)$	-0.610	0.620	1.524	1.297	0.851	57.289
First differences						
$\Delta\log A_t$	-6.281***	-3.204***	-10.513**	-2.279**	0.217**	2.383***
$\Delta\log L_{At}$	-6.387***	-6.337***	-15.712***	-2.802***	0.178**	1.561***
$\Delta\log L_t$	-3.768***	-2.788***	-9.612**	-2.165**	0.225**	2.652**
$\Delta\log(L_{At}/L_t)$	-6.127***	-5.321***	-15.704***	-2.799***	0.178***	1.573***

Notes: The critical values are as follows: (1) *ADF*: -3.646 [-3.661] (1% significance level), -2.954 [-2.960] (5% significance level), -2.616 [-2.619] (10% significance level). (2) *DF–GLS*: -2.637 [-2.642] (1% significance level), -1.951 [-1.952] (5% significance level), -1.611 [-1.610] (10% significance level). (3) MZ_{α}^{GLS}: -13.800 (1% significance level), -8.100 (5% significance level), -5.700 (10% significance level). (4) MZ_{t}^{GLS}: -2.580 (1% significance level), -1.980 (5% significance level), -1.620 (10% significance level). (5) MSB^{GLS}: 0.174 (1% significance level), 0.233 (5% significance level), 0.275 (10% significance level). (6) MP_{T}^{GLS}: 1.780 (1% significance level), 3.170 (5% significance level), 4.450 (10% significance level). The ADF and *DF–GLS* critical values were obtained from the response surfaces of MacKinnon (1996). These critical values are for a (realized) sample size of T = 33 [31]. The critical values for the MZ_{α}^{GLS}, MZ_{t}^{GLS}, MSB^{GLS}, and MP_{T}^{GLS} tests are from Ng and Perron (2001). All unit root tests include a constant. Based on the simulation results of Hall (1994), the lag length for the ADF test was chosen by the general-to-specific method (with a maximum of four lags allowed). For all other tests, we used the modified Akaike information criterion (MAIC) of Ng and Perron (2001). Following the recommendation of Perron and Qu (2006), we calculated the MAIC based on OLS (rather than GLS) detrended data. *** (**) indicate significance at the 1% (5%) level.

Table 3 Cross-sectional dependence and panel unit root tests

		Pesaran's (2007) CIPS panel unit root test			
	CD	$k = 0$	$k = 1$	$k = 2$	$k = 3$
Levels					
$\log A_{it}$	0.000***	0.287	0.103	0.199	0.173
$\log L_{Ait}$	0.000***	0.859	0.970	0.929	0.826
$\log L_{it}$	0.000***	0.447	0.228	0.431	0.509
$\log(L_{Ait}/L_{it})$	0.000***	0.490	0.881	0.199	0.136
First differences					
$\Delta\log A_{it}$	0.000***	0.000***	0.000***	0.011**	0.082*
$\Delta\log L_{Ait}$	0.000***	0.000***	0.000***	0.030**	0.090*
$\Delta\log L_{it}$	0.000***	0.000***	0.001***	0.000***	0.098*
$\Delta\log(L_{Ait}/L_{it})$	0.007***	0.000***	0.000***	0.030**	0.094*

Notes: Reported values are p-values. CD is the cross-sectional dependence test of Pesaran (2004) (adjusted for unbalanced panel data); the null hypothesis is cross-sectional independence. k is the number of lags in the CIPS tests. All tests include country-specific intercepts. *** (**) [*] indicate significance at the 1% (5%) [10%] level.

Table 4 Time-series cointegration tests for the United States

Panel A: Tests for cointegration between $\log L_{At}$ and $\log L_t$ – cointegration tests of Schumpeterian growth models			
	Engle-Granger (1987)	Phillips-Ouliaris (1990)	Pesaran et al. (2001)
p-value of the *t*-statistic	0.719	0.609	
p-value of the *z*-statistic	0.614	0.528	
t-statistic			-1.367
F-statistic			1.208
Panel B: Tests for cointegration between $\log A_t$ and $\log L_{At}$ – cointegration tests of semi-endogenous growth models			
	Engle-Granger (1987)	Phillips-Ouliaris (1990)	Pesaran et al. (2001)
p-value of the *t*-statistic	0.018**	0.013**	
p-value of the *z*-statistic	0.019**	0.011**	
t-statistic			-4.542***
F-statistic			10.508***

Notes: In Panel A, the dependent variable (independent variable) in the tests of Engle and Granger (1987) and Phillips and Ouliaris (1990) is $\log L_{At}$ ($\log L_t$); the dependent variable in the test of Pesaran et al. (2001) test is $\Delta\log L_{At}$. In Panel B, the Engle and Granger (1987) and Phillips and Ouliaris (1990) tests are based on a regression of $\log A_t$ on $\log L_{At}$, while the Pesaran et al. (2001) test is based on a regression with $\Delta\log A_t$ as the dependent variable. The number of lags in the Engle and Granger (1987) and Pesaran et al. (2001) tests was determined using the general-to-specific lag selection procedure, with a maximum of four lags allowed. The 10%, 5% and 1% critical value bounds for the *t*-test of Pesaran et al. (2001) are (-2.57, -2.91), (-2.86, -3.22) and (-3.43, -3.82), respectively. The 10%, 5% and 1% critical value bounds for the *F*-test of Pesaran et al. (2001) are (4.04, 4.78), (4.94, 5.73) and (6.84, 7.84), respectively. The critical value bounds are from Pesaran et al. (2001). If the calculated statistic is above the upper critical value, the null hypothesis of no cointegration can be rejected irrespective. If the calculated statistic is below the lower critical value, the null hypothesis of no cointegration cannot be rejected. If the calculated statistic falls between the lower and upper critical values, the result is inconclusive. *** (**) [*] indicate significance at the 1% (5%) [10%] level.

Table 5 Panel cointegration tests

Panel A: Tests for cointegration between $\log L_{Ait}$ and $\log L_{it}$ – cointegration tests of Schumpeterian growth models			
	Pedroni (1999)		Gengenbach et al. (2016)
	Panel statistics	Group mean statistics	
PP *t*-statistics	-0.197	0.582	
ADF *t*-statistics	-0.059	0.169	
ECM *t*-statistic			-1.948
Panel B: Tests for cointegration between $\log A_{it}$ and $\log L_{Ait}$ – cointegration tests of semi-endogenous growth models			
	Pedroni (1999)		Gengenbach et al. (2016)
	Panel statistics	Group mean statistics	
PP *t*-statistics	-2.764***	-3.438***	
ADF *t*-statistics	-2.404***	-3.665***	
ECM *t*-statistic			-2.935***

Notes: In Panel A, the dependent variable in the Pedroni (1999) tests is $\log L_{Ait}$; the dependent variable in the test of Gengenbach et al. (2016) is $\Delta\log L_{Ait}$. In Panel B, the dependent variable in the Pedroni (1999) tests is $\log A_{it}$; the dependent variable in the tests of Gengenbach et al. (2016) is $\Delta\log A_{it}$. For the Pedroni (1999) (PP and ADF) tests, the lag length was chosen using the general-to-specific method (with a maximum of four lags allowed). The Pedroni (1999) test statistics are distributed as standard normal. The panel variance ratio test has a one-sided rejection region consisting of large positive values, whereas all other tests reject for large negative values. Given the limited number of time-series observations available (for some countries) here, no lags of the first differences were included in the Gengenbach et al. (2016) tests. The critical values for the Gengenbach et al. (2016) *t*-test (for N = 20) are as follows: -2.796 (1% significance level), -2.653 (5% significance level), -1.568 (10% significance level). The critical values are from the online appendix of Gengenbach et al. (2016) (available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2475). To account for cross-sectional dependence (due to possible non-stationary common factors), the results of the Pedroni (1999) tests are based on demeaned data. The Gengenbach et al. (2016) test accounts for cross-sectional dependence via the use of cross-sectional averages. *** (**) indicate significance at the 1% (5%) level.

Table 6 Time-series estimates for the United States of the cointegrating relationship between logA_t and logL_{At}

	DOLS	FMOLS
logL_{At}	0.374***	0.380***
	(0.012)	(0.011)
L_c [p-value]	0.023 [0.693]	0.140 [0.387]
Number of obs.	31	33

Notes: The dependent variable is logA_t. DOLS = dynamic OLS estimator of Stock and Watson (1993); FMOLS = fully modified ordinary least squares estimator of Phillips and Hansen (1990). All regressions include a constant. The DOLS regression was estimated with one lead and one lag. L_c is Hansen's (1992) L_c test for parameter instability in cointegrated relationships; this test is also a test of the null of cointegration against the alternative of no cointegration; the p-values for the L_c test were calculated using the p-value function given by Hansen (1992). Heteroskedasticity and autocorrelation consistent standard errors are in parentheses. *** indicate significance at the 1% level.

Table 7 Panel estimates of the cointegrating relationship between logA_{it} and logL_{Ait}

	PDOLS	PFMOLS
logL_{Ait}	0.168***	0.155***
	(0.025)	(0.024)
p-value of the CD statistic	0.744	0.514
Number of obs.	443	481

Notes: The dependent variable is logA_{it}. PDOLS = panel DOLS estimator of Kao and Chiang (2000); PFMOLS = panel FMOLS estimator of Kao and Chiang (2000). The DOLS regression was estimated with one lead and one lag. All regressions include country fixed effects. The estimators were computed using demeaned data to account for cross-sectional dependence. CD is the cross-sectional dependence test of Pesaran (2004) (adjusted for unbalanced panel data); the null hypothesis is cross-sectional independence. Heteroskedasticity and autocorrelation consistent standard errors are in parentheses. *** indicates significance at the 1% level.

Table 8 Error-correction models for the United States

	Regression with ec_{1t-1}	Regression with ec_{2t-1}
Estimated coefficient on $ec_{1t-1} = \log A_{t-1} - \widehat{\frac{\lambda}{1-\phi}} \log L_{At-1}$, $-\widehat{(1-\phi)}$	-0.344***	
	(0.091)	
Implied ϕ (= $1 - \widehat{(1-\phi)}$)	0.656***	
	(0.091)	
Estimated coefficient on $\Delta\log L_{At-2}$	0.055*	0.055*
	(0.029)	(0.029)
Estimated coefficient on $ec_{2t-1} = \log L_{At-1} - (1/\widehat{\frac{\lambda}{1-\phi}}) \log A_{t-1}$, $\hat{\lambda}$		0.129***
		(0.034)
Number of obs.	31	31

Notes: The dependent variable is $\Delta\log A_t$. ec_{t1} is the residual from the long-run relationship between $\log A_t$ and $\log X_t$, estimated by DOLS; ec_{t2} is the residual from the reverse cointegrating relationship between $\log A_t$ and $\log X_t$, calculated using the reciprocal of the DOLS estimate of the cointegrating relationship. The regressions include a constant. The number of lags in the error-correction regressions was determined using the general-to-specific lag selection procedure, with a maximum of four lags considered. Heteroskedasticity and autocorrelation consistent standard errors are in parentheses. *** (*) indicate significance at the 1% (10%) level.

Table 9 Panel error-correction models

	Regression with ec_{1it-1}	Regression with ec_{2it-1}
Estimated coefficient on $ec_{1it-1} = \log A_{it-1} - \widehat{\frac{\lambda}{1-\phi}} \log L_{Ait-1}$, $-\widehat{(1-\phi)}$	-0.070***	
	(0.011)	
Implied ϕ (= $1 - \widehat{(1-\phi)}$)	0.930***	
	(0.011)	
Estimated coefficient on $\Delta\log L_{Ait}$	0.041***	0.041***
	(0.015)	(0.015)
Estimated coefficient on $ec_{2it-1} = \log L_{Ait-1} - (1/\widehat{\frac{\lambda}{1-\phi}}) \log A_{it-1}$, $\hat{\lambda}$		0.012***
		(0.002)
p-value of the CD statistic	0.593	0.593
Number of obs.	481	481

Notes: The dependent variable is $\Delta\log A_{it}$. ec_{it1} is the residual from the long-run relationship between $\log A_t$ and $\log X_t$, estimated by DOLS; ec_{it2} is the residual from the reverse cointegrating relationship between $\log A_t$ and $\log X_t$, calculated using the reciprocal of the panel DOLS estimate of the cointegrating relationship. All regressions include country fixed effects. The estimates control for error cross-sectional dependence via the use of (weighted) cross-sectional averages, following the common correlated effects approach of Pesaran (2006). CD is the cross-sectional dependence test of Pesaran (2004) (adjusted for unbalanced panel data); the null hypothesis is cross-sectional independence. We used the common recursive mean adjustment to reduce the dynamic panel bias. Heteroskedasticity and autocorrelation consistent standard errors are in parentheses. *** indicate significance at the 1% level.

Appendix

Table A1 Time-series unit root tests for the United States for the log of patent applications (logPat_t) and the log of the patent stock (log$PatStock_t$), 1980-2014

	ADF	*DF–GLS*	MZ_{α}^{GLS}	MZ_t^{GLS}	MSB^{GLS}	MP_T^{GLS}
logPat_{it}	-0.554	0.438	1.002	0.911	0.909	58.522
ΔlogPat_{it}	-5.913***	-5.641***	-15.954***	-2.805***	0.176**	1.609***
log$PatStock_{it}$	-0.226	-0.968	1.832	3.501	1.911	276.707
Δlog$PatStock_{it}$	-2.144	-1.517	-3.130	-1.234	0.394	7.798
Δ^2log$PatStock_{it}$	-6.128***	-5.459***	-15.382***	-2.757***	0.179**	1.654***

Notes: The critical values are as follows: (1) *ADF*: -3.646 [-3.661] (1% significance level), -2.954 [-2.960] (5% significance level), -2.616 [-2.619] (10% significance level). (2) *DF–GLS*: -2.637 [-2.642] (1% significance level), -1.951 [-1.952] (5% significance level), -1.611 [-1.610] (10% significance level). (3) MZ_{α}^{GLS}: -13.800 (1% significance level), -8.100 (5% significance level), -5.700 (10% significance level). (4) MZ_t^{GLS}: -2.580 (1% significance level), -1.980 (5% significance level), -1.620 (10% significance level). (5) MSB^{GLS}: 0.174 (1% significance level), 0.233 (5% significance level), 0.275 (10% significance level). (6) MP_T^{GLS}: 1.780 (1% significance level), 3.170 (5% significance level), 4.450 (10% significance level). The ADF and *DF–GLS* critical values were obtained from the response surfaces of MacKinnon (1996). These critical values are for a (realized) sample size of T = 33 [31]. The critical values for the MZ_{α}^{GLS}, MZ_t^{GLS}, MSB^{GLS}, and MP_T^{GLS} tests are from Ng and Perron (2001). All unit root tests include a constant. Based on the simulation results of Hall (1994), the lag length for the ADF test was chosen by the general-to-specific method (with a maximum of four lags allowed). For all other tests, we used the modified Akaike information criterion (MAIC) of Ng and Perron (2001). Following the recommendation of Perron and Qu (2006), we calculated the MAIC based on OLS (rather than GLS) detrended data. *** (**) indicate significance at the 1% (5%) level.

Table A2 Cross-sectional dependence and panel unit root tests for the log of patent applications (logPat_{it}) and the log of the patent stock (log$PatStock_{it}$), 1980-2014

		Pesaran's (2007) CIPS panel unit root test			
	CD	$k = 0$	$k = 1$	$k = 2$	$k = 3$
logPat_{it}	0.000***	0.700	0.812	0.896	0.989
ΔlogPat_{it}	0.944	0.000***	0.000***	0.001***	0.064*
log$PatStock_{it}$	0.000***	0.726	0.130	0.591	0.995
Δlog$PatStock_{it}$	0.000***	0.954	0.682	0.906	0.995
Δ^2log$PatStock_{it}$	0.000***	0.000***	0.000***	0.070*	0.889

Notes: Reported values are p-values. CD is the cross-sectional dependence test of Pesaran (2004) (adjusted for unbalanced panel data); the null hypothesis is cross-sectional independence. k is the number of lags in the CIPS tests. All tests include country-specific intercepts. *** (*) indicate significance at the 1% (10%) level.