Economic Growth and Innovation Complexity: An Empirical Estimation of a Hidden Markov Model

Alberto Bucci* Lorenzo Carbonari† Pedro Mazeda Gil‡

Giovanni Trovato§

Abstract

Over the past decades, research effort in high income countries has substantially increased. Meanwhile, the growth rates of per capita *output have been rather stable. The contribution of this paper is twofold. The first is to provide a theoretical explanation for such trends by developing an R&D-based growth model which accounts for dilution, difficulty and duplication effects. The second is to show empirically that the occurrence of different phases in the economic growth dynamics traces back to the interplay between complexity and specialization in production. To do this we estimate a Hidden Markov Model in which countries can switch across different growth regimes. We identify four distinct growth regimes.*

Keywords: Economic growth; Population growth; Complexity; Hidden Markov Model.

JEL codes: O3; O4; J1.

*Università di Milano (DEMM and FinGro Lab), CEIS Tor Vergata, and RCEA (Rimini Center for Economic Analysis). E-mail address: alberto.bucci@unimi.it.

†Università di Roma "Tor Vergata", DEF and CEIS. Corresponding author. E-mail address: lorenzo.carbonari@uniroma2.it.

‡Universidade do Porto and CEF.UP. E-mail address: pgil@fep.up.pt.

§Università di Roma "Tor Vergata", DEF and CEIS. E-mail address: giovanni.trovato@uniroma2.it.

"*Rich countries specialize in complicated products*".

Kremer (1993), p. 563.

1 Introduction

Modern growth literature since Romer (1990) places purposeful innovation and R&D activities at the center of the growth process. Yet, empirically, it is apparent the non-significant or even negative empirical correlation between R&D effort and the growth rates of *per capita* ouput in developed countries. This paper contributes with a theoretical explanation for such trends by developing an R&D-based growth model which encompasses both specialization benefits and complexity costs in production and R&D activities. Then, we take advantage of this setting to explore, empirically, the link between the occurrence of different growth regimes and the variation of the specialization and complexity effects across time and countries. Figures 1 and 2 provide some aggregate evidence, from a group of advanced economies over the period 1983-2007, concerning the relationship between long-run *per capita* real GDP growth and innovation, captured by either research inputs (e.g., number of researchers employed per million inhabitants) or research outputs (e.g., number of patent applications per resident). With the partial exception of the UK, no positive correlation seems to exist between long-run *per capita* real GDP growth and innovation, captured by either research inputs (e.g. number of researchers employed per million inhabitants) or research outputs (e.g. number of patent applications by residents). Data from other OECD countries or other measures of output (e.g. the real GDP per hours worked) would produce similar figures. Of course, one may question the economic and statistical significance of this suggestive evidence, since aggregate innovation is a broad phenomenon, only partially captured by these two R&D related measures. Nonetheless, the increasing trends reported in Table 1, in terms of patenting activity and researchers employed, have, at least, no corresponding increase in the long-run growth rates of *per* *capita* GDP. Long-run income and productivity growth have rather declined over the last twenty years (apart from the modest recovery immediately before the 2007-2008 global financial crisis).[1]

Table 1: Four OECD countries, 1983-2007

Country	5Y-AVG. Real GDP growth (% variation)	Patent appl. by residents[a] (% variation)	Researchers[b] (% variation)
Germany	+1.10	+51.45	+107.69
Japan	-2.03	+11.63	+46.70
United Kingdom	-1.10	-12.88	+83.10
United States	-1.50	+253.28	+52.96

Note: Japanese data are referred to the period 1983-2006. Source: [a] World Bank, [b] (per mil. of inhab.) OECD.

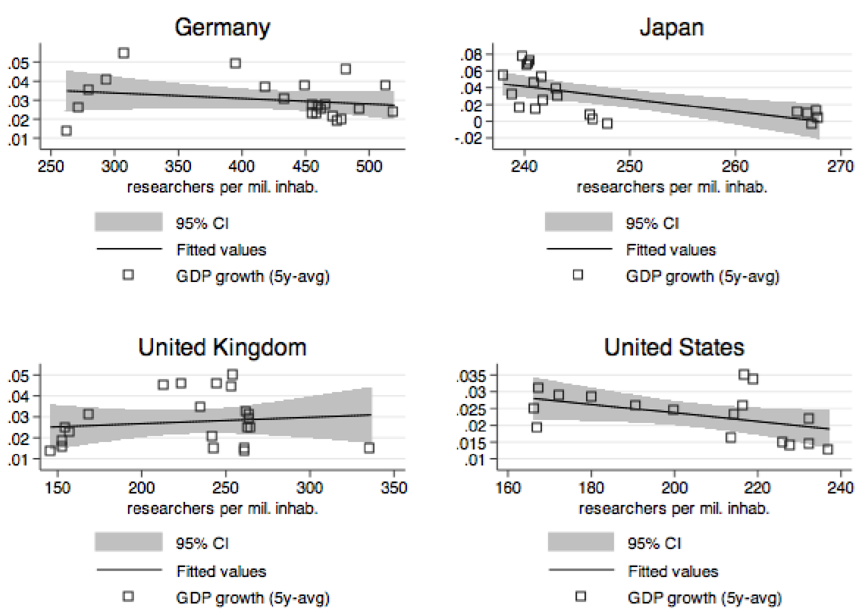

Figure 1: Real GDP growth and Research Input

The main contribution of this paper is to provide a possible explanation for such diverging trends. In the spirit of Aghion and Howitt (1998), our explanation is based on the idea that innovation makes production activities more complex and this requires

[1] Similar empirical results have been obtained by, e.g., Backus, Kehoe and Kehoe (1992), Bassanini, Scarpetta and Visco (2000), Pintea and Thompson (2007), and Gil, Brito and Afonso (2013), either from a cross-sectional or a time-series perspective.

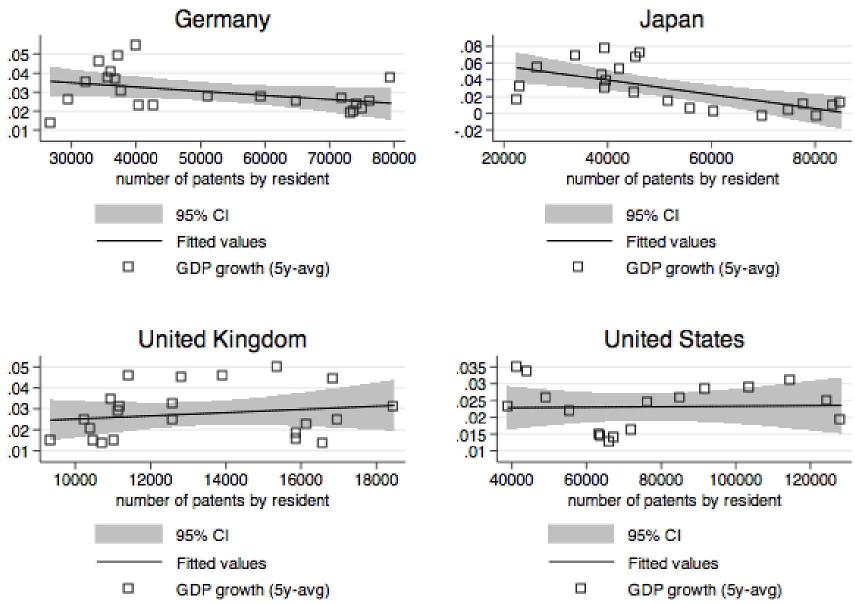

Figure 2: Real GDP growth and Research Output

the development of appropriate skills and abilities to adapt to changing technological needs. We develop an R&D-based endogenous growth model in which a larger number of intermediate-input varieties to be combined within the same manufacturing process simultaneously causes an increase in the cost, due to the higher complexity, and new opportunities of gain, originating from specialization. Complementarily, we explore the idea, in the spirit of Jones (1995) and Ha and Howitt (2007), that innovation benefits from increasing research inputs but may be curtailed by complexity effects pertaining to innovation activities (under the form of duplication, difficulty and dilution effects), as the number of varieties of intermediates and the amount of research inputs rise in the economy.

The second contribution of the paper is to show empirically that the occurrence of different phases in the economic growth dynamics traces back to the interplay between complexity and specialization in production. The tension between complexity (both in production and innovation activities) and specialization relies on the mechanism through which population growth and monopolistic markups may at the same time impact economic growth. To rationalize this mechanism, we propose an extended version of Jones (1995) *semi-endogenous* growth model. To allow for a variety of growth phases, we assume – in the econometric part of the paper – that there exist different "states of nature", in which growth behavior (captured by some key structural parameters) differs for otherwise identical countries.[2] To identify the growth regimes and the transition between them, we estimate our theoretical growth equation using a Hidden Markov Model, which allows to deal with both observed and unobserved (*hidden*) factors that affect long-run growth (Baum and Petrie, 1966; Baum and Eagon, 1967). For a sample of OECD countries, we find four distinct regimes, corresponding to four growth processes. These regimes capture different balanced growth paths, with different long-run average growth and different growth volatility. Importantly, our results are robust to several sensitivity checks and alternative estimation techniques. In particular, our growth regimes classification survives when we employ an alternative theoretical specification in which we (partially) endogenize complexity by letting some of its determinants be functions of the research input (i.e., researchers/population).

The gist of our argument is the following. R&D activities may generate conflicting effects on long-run growth. Hence, depending on the parameter space, the R&D intensity can produce either a faster or a slower real GDP growth along the balanced growth path. In particular, the specialization effect tends to prevail on the complexity effect when the (constant) markup charged by the intermediate-good firms on final-good producers is not too low. In the empirical part of the paper, we allow these parameters to vary across time and across countries and we identify different growth regimes for our sample of OECD economies. Countries switch between different regimes. Our estimates confirm that the production of new ideas, which originates from the interaction between patents and research (within the R&D production function), is the engine of growth. Typically, higher long-run growth rates are found associated to higher saving

[2]In this sense, the regime-varying parameters are reduced forms of the underlying processes generating aggregate complexity. Hence, (possible) changes in these parameters may be interpreted as a change in the complexity generated by these processes.

and investment rates and to lower levels of public expenditure to GDP ratio. Moreover, for those countries stuck in a low growth regime, improving the overall education attainment increases the probability to move towards higher growth regimes.

After discussing the related literature (Section 2), we develop our theoretical model (Section 3) and then present our econometric analysis (Section 4). Section 5 provides concluding remarks and policy implications.

2 Related literature

From the theoretical standpoint, the paper closest to ours is Bucci, Carbonari and Trovato (2019), who analyze how production complexity may affect the long-run GDP growth rate and its correlation with population growth and the intermediate sector's markups. The model presented below extends their R&D technology to simultaneously deal with the difficulty, duplication and dilution effects in innovation activities. Moreover, unlike Bucci, Carbonari and Trovato (2019), in the empirical part of the paper, we use the restrictions provided by the model to analyze, from an econometric standpoint, how complexity may affect not only the transition between different phases of economic development but also the membership to a specific growth regime.

Other works related to our theoretical framework also include Aghion and Howitt (1998, ch. 12), Dinopoulos and Thompson (1998), Howitt (1999), Peretto (1998), Peretto and Smulders (2002), and Young (1998). These papers represent another (in this case, fully-endogenous growth) class of second-generation R&D-based growth models (see Jones, 1999) and, like us, are able to nicely capture a dilution effect in R&D activities. However, differently from the semi-endogenous growth tradition (Jones, 1995; Kortum, 1997; Segerstrom, 1998) that directly inspires our theory, the models of Aghion and Howitt (1998, ch. 12), Dinopoulos and Thompson (1998), Howitt (1999), Peretto (1998), Peretto and Smulders (2002), and Young (1998) formalize the idea that as an economy grows, the rise of product-variety (which is proportional to the population in the economy) reduces the R&D effort directed to quality upgrading and ultimately causes research-investments to be spread more thinly over a larger number of different varieties of goods. In other words, while this class of 'Schumpeterian' growth models copes with the dilution effect by invoking the simultaneous presence of the two canonical dimensions of R&D (i.e., the horizontal and vertical ones, respectively), in the present paper we focus solely on the variety-dimension of innovation. Following Ha and Howitt, 2007, we provide below an economic justification of how a dilution-effect, due to population size, can arise within a framework where innovation is solely of the horizontal type.

In parallel, there is an empirical debate on the semi-endogenous versus fully-endogenous growth framework within the second-generation R&D-based growth literature, which our paper also contributes to – see, e.g., Laincz and Peretto (2006), Ha and Howitt (2007), Madsen (2008, 2010), Madsen and Ang (2011), and Ang and Madsen (2015). We use a different empirical methodology from all these papers, which allows us to obtain country-specific and regime-specific measures of the difficulty, duplication, and dilution effects. In all cases, we detect the existence of difficulty effects in line with the semi-endogenous growth approach, whereas the estimated dilution effects do not, in general, support the fully-endogenous growth approach.

Other papers have recently addressed the link between complexity and growth. Pintea and Thompson (2007) is a significant contribution. However, with respect to ours, they use a radically different approach that allows them to identify in the increasing technological complexity, experienced during the second half of 20th century, the main cause for the puzzling coexistence of secular increases in R&D expenditure and educational attainment and no corresponding increase in *per capita* income growth.

Using trade data and the notion of "product space", Hausmann and Klinger (2006) develop a model offering a broad appraisal of the complexity effect. Along this line and using a sample of 89 countries over the period 1990-2009, Ferrarini and Scaramozzino

(2012) support the idea that, mainly for high income countries, gains from specialization are smaller than the corresponding losses due to increased complexity. Differently from them, we study how the balance between specialization and complexity influences both the long-run GDP growth rate and the growth regime in which a country may fall.

Finally, within the literature which reads economic growth as a sequence of transitions between distinct growth phases that countries visit with different frequencies, two contributions deserve to be mentioned, namely Jerzmanowski (2006) and Kerekes (2012). Both papers deliver purely empirical exercises: in particular, Jerzmanowski (2006) studies 89 countries over the period 1962-1994 while Kerekes (2012) studies 84 countries over the period 1961-2003. They employ an econometric technique – namely, a Markow switching model – similar to ours. Differently from us, in both papers growth dynamics simply originates from an auto-regressive process and there is no room for the implications of complexity in determining the long-run pattern of the real *per capita* GDP.

3 The Model

We build upon the R&D-driven growth model of Romer (1990) and the extensions of this model provided, respectively, by Grossman and Helpman (1991) and Bucci, Carbonari and Trovato (2019). These are dynamic general-equilibrium endogenous growth models where a homogeneous final (consumption) good is competitively produced using labor and a continuum of varieties of intermediate inputs. The latter are, in turn, produced in a monopolistically-competitive sector using solely (one-to-one) labor. Potential entrants into the intermediate-good sector devote labor to horizontal R&D, by which they increase the measure of varieties of intermediate inputs. The economy is populated by infinitely-lived (dynastic) households who consume and inelastically supply labor to firms in the final-good, the intermediate-good, and the R&D sectors. We assume that the aggregate labor force coincides with total population, which increases at a constant positive exogenous growth rate.

3.1 Production

The final good, Y, is produced competitively at time t using, as private and rival inputs, both labor and a continuum of intermediate goods,

$$Y_t = L_{Yt}^{1-\alpha} \left[\frac{1}{N_t^{\beta}} \int_0^{N_t} (x_{it})^{1/m} di \right]^{\alpha m}, \qquad 0 < \alpha < 1 \quad \text{and} \quad m > 1. \tag{1}$$

In the equation above, L_{Yt} is the quantity of the labor input in final-good production, x_{it} is the quantity of the *i-th* variety of differentiated intermediate inputs with $i \in [0, N_t]$, α is a parameter that controls for the the labor share in final-good production (this share is given by $1 - \alpha$), and m is a parameter that controls for the elasticity of substitution between any generic pair of varieties of intermediate goods (this elasticity is given by $m/(m-1)$). Following Ethier (1982) and Benassy (1996a, 1996b, 1998), the aggregate production function (1) allows to disentangle the optimal markup on the marginal production cost in the intermediate-good sector (or, alternatively, the measure of product market concentration in that sector), m, from the factor-shares in final-good production, α and 1-α. As we show below, a decrease in m increases the substitutability between intermediate goods and, thus, leads to tougher competition across intermediate-good firms and to lower prices. Therefore, m can be regarded as a (inverse) measure of the degree of competition in the intermediate-good market.[3] Finally, parameter β controls for the *complexity effect* on aggregate output induced by the expansion of varieties of intermediate goods, as emphazised by, e.g., Aghion and Howitt (1998). In particular when positive, β captures the deleterious effect on aggregate output of having a larger number of intermediate-input varieties to be combined in the same productive process. This effect contrasts with the standard

[3] See, e.g., Bucci (2013), Ethier (1982) and Benassy (1996a, 1996b, 1998).

positive specialization effect which results from the increasing availability of differentiated intermediate inputs in aggregate production and which is represented by the time-varying upper bound of the integral in equation (1). Assuming symmetry, i.e. $x_{it} \equiv x_t > 0 \ \forall i \in [0, N_t]$, and with $L_{Yt} > 0$ and $N_t \in (0, \infty)$, equation (1) indicates that an increase in N may have either a net positive (if $\beta < 1$), a net negative ($\beta > 1$) or a null net impact on aggregate output ($\beta = 1$). Notice that $\beta < 0$ implies that there is no *complexity effect* (the expansion in variety of intermediate goods would, in this case, amplify the standard specialization effect referred to above) whereas $\beta = 0$ trivially implies that only the standard specialization effect exists. Hence, to model explicitly a complexity effect due to an increase in N, our model needs some positive β. However, no *ad hoc* assumption is hereby made on the sign and the magnitude of β. Given perfect competition in the final-good sector, producers take wages, w_{Yt}, and input prices, p_{it}, as given and sell their output at a price also taken as given (and which we normalize to unity).

Each intermediate good, x_i, is produced in a monopolistically-competitive sector using labor as the sole input. The sector uses the technology (see Grossman and Helpman, 1991):

$$x_{it} = l_{it}, \ \forall i \in [0, N_t], \ \ N_t \in [0, \infty), \tag{2}$$

where l_{it} is the labor necessary to produce the *i-th* intermediate good. Given this technology, the wage rate w_{xt} is the marginal cost of production. For a given N_t, equation (2) implies that the total amount of labor employed in the intermediate-good sector at time t, L_{xt}, is given by $\int_0^{N_t} x_{it} di = \int_0^{N_t} l_{it} di = L_{xt}$. Maximization of the *i-th* firm's instantaneous profit leads to the standard markup rule $p_{it} \equiv p_t = m w_{xt}, \ \forall i \in [0; N_t]$.[4] Because of the symmetry of producers of intermediate goods, the price is

[4] We assume that each of these firms is so small to take $\left[\frac{1}{N_t^\beta}\int_0^{N_t}(x_{it})^{1/m}di\right]^{\alpha m-1}$ as given, so that $\frac{\partial}{\partial x_{it}}\left[\frac{1}{N_t^\beta}\int_0^{N_t}(x_{it})^{1/m}di\right]^{\alpha m-1} = 0$

homogeneous across all intermediate goods i and equal to a constant markup, m, on the marginal production cost, w_{xt}. This then implies $x_{it} \equiv x_t = L_{xt}/N_t$, $\forall i \in [0; N_t]$, and, thus:

$$\pi_{it} \equiv \pi_t = \alpha \left(\frac{m-1}{m}\right) \left(\frac{L_{Yt}}{N_t}\right)^{1-\alpha} \left(\frac{L_{xt}}{N_t}\right)^{\alpha} N_t^{\alpha[m(1-\beta)-1]}, \ \forall i \in [0; N_t]. \tag{3}$$

3.2 R&D sector

R&D is performed by (potential) entrants in the intermediate-good sector. A successful innovation leads to a new blueprint pertaining to a new variety of intermediate good, which is granted a perpetual patent. The R&D sector is characterized by i) free entry, ii) perfect competition and iii) constant returns to scale. At the aggregate level, the R&D production function is given by:

$$\begin{aligned} \dot{N}_t &= \frac{1}{X} \cdot \underbrace{N_t}_{\text{knowledge spillover}} \cdot \underbrace{\frac{1}{N_t^{\chi_1}}}_{\text{difficulty effect}} \cdot \underbrace{\frac{1}{L_t^{\chi_2}}}_{\text{dilution effect}} \cdot \underbrace{\frac{1}{L_{Nt}^{1-\lambda}}}_{\text{duplication effect}} \cdot L_{Nt} \\ &= \frac{1}{X} \cdot N_t^{1-\chi_1} \cdot L_t^{-\chi_2} \cdot L_{Nt}^{\lambda}, \end{aligned} \tag{4}$$

where N_t is the number of already invented ideas, L_{Nt} is the labor input in the R&D sector, L_t is the total labor force in the economy (and which works as a scale variable for the market dimension), and $X > 0$ is a parameter that controls for the efficiency in R&D activities.

The component $\bar{\delta} \equiv \frac{1}{X} \cdot N_t^{1-\chi_1} \cdot L_t^{-\chi_2} \cdot L_{Nt}^{\lambda-1}$ in equation (4) is external to each individual R&D firm/researcher. Within this completely external component, N_t captures (as in Romer, 1990) the presence of a positive intertemporal knowledge-spillover effect. When $\chi_1 > 0$, the term $N_t^{-\chi_1}$ accounts, instead, for the occurrence of a difficulty effect in innovation (see, e.g., Jones, 1995 and Segerstrom, 1998, among others).[5] When

[5] The difficulty effect in innovation captures the notion that ideas that are easier to discover tend to be

$\chi_2 > 0$, the term $L_t^{-\chi_2}$ denotes the canonical R&D dilution-effect due to population size (as, e.g., in Ha and Howitt, 2007)[6]. Finally, when $0 < \lambda \leq 1$, the term $L_{Nt}^{\lambda-1}$ describes (as in Jones, 1995) the possible duplication-effect related to the amount of labor input employed in R&D. Equation (4) represents the main departure from Bucci, Carbonari and Trovato (2019) who, in its place, employ an otherwise standard Jones (1995)-type aggregate ideas production function.

The parameters χ_i, $i = 1, 2$, may be interpreted as (factor-specific) complexity indices pertaining to R&D activity, as in Sequeira, Gil and Afonso (2018). Notice that these complexity factors add to the complexity effect we consider in final-good production, as captured by a positive value of β in equation (1).

When $\chi_1 > 0$, higher values of N_t imply that the same amount of R&D resources (L_{Nt}) generates a lower rate of innovation, $\dot{N}_t/N_t$, i.e., there are decreasing technological opportunities. This is crucial to eliminate the strong scale effect on growth found in the first-generation of R&D-based growth models (e.g., Jones, 1995): since the marginal impact of an individual researcher on the growth rate of new ideas decreases with the stock of existing ideas, it is possible to sustain a constant positive rate of innovation only by increasing (at a constant rate, too) the number of researchers. This, in turn, is possible in long-run equilibrium solely if population grows at a positive rate. On the other hand, $\chi_2 = 1$ also allows for the removal of the strong scale effect, even if $\chi_1 = 0$. In this case, it is possible to maintain a constant positive rate of innovation as long as the proportion of the number of researchers in the population, L_{Nt}/L_t, is constant. In this scenario, a positive rate of innovation does not require that population grows at a positive rate.

To sum up, the R&D function (4) features a general formulation in line with Ha and Howitt (2007), and nests the following specific well-known cases: under $\chi_1 = \chi_2 = 0$

discovered first, making it harder to find new ideas subsequently.

[6]An economic interpretation of this effect is that "*... a larger population increases the number of people who can enter an industry with a new product, thus resulting in more horizontal innovations, which dilutes R&D expenditure over a larger number of separate projects*" (Ha and Howitt, 2007).

(and $\lambda = 1$), we recover the fully endogenous growth model with strong scale effect by Romer (1990); under $\chi_1 > 0$, and $\chi_2 = 0$ (and $\lambda \leq 1$), we get the *semi-endogenous* growth model without the strong scale effect by Jones (1995); under $\chi_1 = 0$, and $\chi_2 = \lambda$, we get the fully endogenous growth model without the strong scale effect, as e.g. in Dinopoulos and Thompson (1999, 2000).

Since the R&D sector is competitive, N_t is endogenously determined so that the wage rate of one unit of research labor input satisfies the free-entry condition $w_{Nt}L_{Nt} = \dot{N}_t V_{Nt}$. This condition suggests that entry into the R&D sector will cease when total revenues from the discovery of new ideas (the RHS) equal total direct costs related to ideas-production (the LHS), with V_{Nt} indicating the market value of any new idea being discovered.

3.3 Households

We normalize the number of the identical infinitely-lived households to one, so that the population/labor force size coincides with that of the single dynastic family, L, which supplies labor inelastically.

Population grows at a constant exogenous rate $\dot{L}_t/L_t \equiv n > 0$. In this model, as shown above, labor is employed for the production of final (consumption) goods (L_Y), intermediate goods (L_x), and ideas (L_N).

$$L_t = L_{Yt} + L_{xt} + L_{Nt} \tag{5}$$

Since we assume that labor is homogeneous and perfectly mobile, at equilibrium, wages must equalize:

$$w_{Yt} = w_{xt} = w_{Nt} \equiv w_t \tag{6}$$

The representative dynastic family solves:

$$\max_{\{c_t,a_t\}_{t=0}^{\infty}} U \quad \equiv \quad \int_0^{\infty} \left(\frac{c_t^{1-\theta}-1}{1-\theta} \right) e^{-(\rho-n)t} dt, \tag{7}$$

subject to the usual flow budget contraint, $\dot{a}_t = (r_t - n)a_t + w_t - c_t$, and $a_0 > 0$ given, where $a_t \equiv A_t/L_t$ and $c_t \equiv C_t/L_t$ denote *per capita* asset holdings and *per capita* consumption, respectively. In equation (7), $\theta > 0$ is the inverse of the intertemporal elasticity of substitution in consumption, $\rho > 0$ is the pure subjective discount rate and we have normalized population at time 0 to one, $L_0 = 1$. The representative dynastic family chooses the optimal path of *per capita* consumption and asset holdings, $\{c_t, a_t\}_{t=0}^{\infty}$, taking the real rate of return on asset holdings (the real interest rate), r_t, and the wage rate, w_t, as given.

3.4 Equilibrium and Balanced-Growth Path

In this economy, an *allocation* is a set of time paths for *per capita* consumption and asset holdings $\{c_t, a_t\}_{t=0}^{\infty}$, for the available number of intermediate-good varieties $\{N_t\}_{t=0}^{\infty}$, for prices and quantities of each intermediate good $\{p_t, x_t\}_{t=0}^{\infty}$, and for the real interest rate and wages $\{r_t, w_t\}_{t=0}^{\infty}$. We define an *equilibrium* as an allocation in which: i) the time paths for consumption and asset holdings are consistent with the solutions of the households' problem (7); ii) the time paths for prices and quantities of each intermediate good maximize instantaneous profits (3); iii) the time path for the number of intermediate-good varieties is determined by the free-entry condition in the R&D sector; and iv) the time paths for the real interest rate and wages are consistent with market clearing.

We can now characterize the *Balanced-Growth Path* (BGP) of this model. A BGP equilibrium is an equilibrium path along which: i) All variables grow at constant exponential rates and ii) the sectoral shares of labor employment ($s_j = \frac{L_{jt}}{L_t}$ with $j =$

$Y, x, N)$ are constant.

Let $\chi_1 \neq 0$. It can then be shown that, along a BGP equilibrium:[7]

$$\gamma_N \equiv \frac{\dot{N}_t}{N_t} = \Psi n \tag{8}$$

$$\gamma_c \equiv \frac{\dot{c}_t}{c_t} = \gamma_a \equiv \frac{\dot{a}_t}{a_t} = \gamma_y \equiv \frac{\dot{y}_t}{y_t} = \Phi\Psi n \tag{9}$$

$$r = \theta\Phi\Psi n + \rho \tag{10}$$

where:

$$\Phi \equiv \alpha\left[m(1-\beta) - 1\right] \gtreqless 0 \tag{11}$$

and

$$\Psi \equiv \left(\frac{\lambda - \chi_2}{\chi_1}\right) \gtreqless 0 \tag{12}$$

and with the transversality condition of the households' optimisation problem satisfied when $\rho > n + (1-\theta)\gamma_y$. Equation (8) gives the BGP equilibrium growth rate of the number of varieties of intermediate goods in the economy. According to equation (9), the *per capita* values of consumption, asset holdings, and income, $y \equiv Y/L$, grow at the same constant rate along the BGP. Equation (10) gives the BGP value of the real rate of return on asset holdings.

As in any canonical R&D-based growth model, the BGP growth rate of *per capita* income is strictly related to the BGP innovation rate: $\gamma_y = \Phi\gamma_N$. At the same time, as in the basic *semi-endogenous growth* model (Jones, 1995), and provided that $\chi_1 \neq 0$, the innovation rate (γ_N) depends only on the parameters of the innovation technology (λ, χ_1, and χ_2) and the exogenous population growth rate (n), while it is independent of the markup (m) and the parameter measuring the degree of complexity in final output production due to input proliferation (β). Although m and β do not affect γ_N in the BGP equilibrium, they do have an impact, respectively, on the market value of

[7] The derivations are available from the authors upon request.

any generic idea (through π_i), and on the wage accruing to one unit of research labor (w_N). As a consequence, m and β are ultimately able, by means of wage equalization, to influence the allocation of the available labor input across sectors (see equations (5) and (6)).

Notice that the BGP growth rate γ_N may display a positive, null, or negative relationship with the population growth rate, n, depending on the sign and magnitude of the innovation technology parameters. Unlike γ_N, through Φ, γ_y depends instead on m and β. This entails that the degree of market power, the overall level of complexity in the economy, and the interactions between these two variables ultimately determine the output growth rate in the long-run and its correlation with n.

From equations (8)-(10), it is evident that although γ_N is positive if $\Psi > 0$, r is positive if $\Phi\Psi$ is sufficiently large, i.e. $\Phi\Psi > -\frac{\rho}{\theta n}$. Instead, for γ_y to be positive, $\Phi\Psi$ needs to be strictly greater than zero. In what follows, we present the results of our model in their most general possible form without imposing any *ex ante* restriction on the magnitude of the upper bound of β.

We can easily show that along the BGP:[8]

1. $Sign(\gamma_y) = Sign(\Phi\Psi)$;
2. $Sign\left(\frac{\partial \gamma_y}{\partial n}\right) = Sign(\Phi\Psi)$;
3. Real *per capita* income growth is equal to zero in the absence of any population change (i.e., when $n = 0$);
4. $Sign\left(\frac{\partial \gamma_y}{\partial m}\right)$ depends simultaneously on $Sign(\Psi)$ and on whether $\beta \gtreqless 1$.

Using the definition of Φ and equation (9), we observe that several combinations of the signs of γ_y, $\frac{\partial \gamma_y}{\partial n}$, and $\frac{\partial \gamma_y}{\partial m}$ may occur when $\Psi > 0$ (say because $\chi_1 > 0$ and $\lambda > \chi_2$ in equation (4)):

- if $0 < \beta < \frac{m-1}{m}$, then: $\Phi > 0$, $\gamma_y > 0$, $\frac{\partial \gamma_y}{\partial n} > 0$ and $\frac{\partial \gamma_y}{\partial m} > 0$;
- if $\beta = \frac{m-1}{m}$, then: $\Phi = 0$, $\gamma_y = 0$, $\frac{\partial \gamma_y}{\partial n} = 0$ and $\frac{\partial \gamma_y}{\partial m} = 0$;

[8]See Appendix A.1 for the analytical derivation.

- if $\frac{m-1}{m} < \beta < 1$, then: $\Phi < 0$, $\gamma_y < 0$, $\frac{\partial \gamma_y}{\partial n} < 0$ and $\frac{\partial \gamma_y}{\partial m} > 0$;
- if $\beta = 1$, then: $\Phi < 0$, $\gamma_y < 0$, $\frac{\partial \gamma_y}{\partial n} < 0$ and $\frac{\partial \gamma_y}{\partial m} = 0$;
- if $\beta > 1$, then: $\Phi < 0$, $\gamma_y < 0$, $\frac{\partial \gamma_y}{\partial n} < 0$ and $\frac{\partial \gamma_y}{\partial m} < 0$.

Our model indicates that the sign of $(1-\beta)$, by determining the sign of $\partial\Phi/\partial m$, is a key determinant of the impact of m on γ_y. In more detail, when $\beta < 1$, the specialization gains originating from an expansion in intermediate-input variety are larger than the potential losses due to more complexity and, thus, $\frac{\partial \Phi}{\partial m} > 0$. When $\beta > 1$, the specialization gains are smaller than the potential losses due to more complexity and $\frac{\partial \Phi}{\partial m} < 0$. Finally, when $\beta = 1$, the specialization gains are offset by the potential losses due to more complexity and $\frac{\partial \Phi}{\partial m} = 0$. These results, combined with the sign of Ψ, determine the sign of $\frac{\partial \gamma_y}{\partial m}$, as shown above. The sign of Ψ depends, in turn, on the sign of $(\lambda - \chi_2)$, that is, on the balance between dilution and duplication effects in R&D activities, as shown by equation (4).

The result that, in the absence of demographic change, there is no growth in real *per capita* income is a distinctive characteristic of any basic *semi-endogenous growth* model (see Jones, 1995). Our setting, however, presents at least three additional features that cannot be found in a canonical *semi-endogenous growth* model. First, here the relation between γ_N and γ_y is mediated by the term $\Phi \equiv \alpha\,[m(1-\beta)-1]$, unlike Jones (1995), where $\gamma_N = \gamma_y = \gamma$, with γ being a definitely positive function of n. Thus, while in Jones (1995), $\beta = 0$ and then $\Phi \equiv \alpha(m-1) > 0$, in our model Φ can also be negative. This happens, for any given $m > 1$, when β is sufficiently large, $\beta > (m-1)/m \in (0,1)$, i.e. when the complexity effect (in final-good production) is strong enough. When this is true ($\Phi < 0$), innovation output and long-run growth may exhibit diverging patterns, as in the suggestive evidence provided at the beginning of the paper. In this case, it is also possible that an increase in n would yield (unlike Jones, 1995) a negative impact on γ_y. This occurs whenever $\Psi > 0$, as n affects positively γ_N in such a case.

Second, in equation (9), γ_y depends not only on n and, among others, the parameters χ_1, χ_1, λ and β, but also on m.

Third, unlike Jones (1995), in our model $\Psi \equiv \left(\frac{\lambda - \chi_2}{\chi_1}\right) \gtreqless 0$, because of our explicit and simultaneous consideration of a dilution, a difficulty, and a duplication effect in R&D activity. Depending on the combined sign of Ψ and Φ, innovation, (physical) inputs and economic growth may exhibit diverging patterns, also as in the suggestive evidence presented at the beginning of the paper.

4 Quantitative analysis

The theory developed above rests on a precise notion of complexity, which emerges as a consequence of the interaction between the entire labor force and the aggregate research effort, i.e. the number of new ideas and of researchers at work. The combination between "complexity parameters"(χ_1, χ_2, λ and β), "technological parameters"(α and m) and the exogenous population growth rate (n) determines the long-run growth rate of a country's *per capita* income. We now confront the theoretical predictions with OECD data, by allowing that these fundamental parameters may vary over time and across countries. To do this, we employ a Hidden Markov Model (HMM, hereafter), which is specified by the following components: i) a set of states (or *regimes* or *classes*), ii) a transition probability matrix, iii) a sequence of observations, iv) a sequence of observation likelihoods and v) an initial probability distribution over states. The main advantage of this econometric technique is that it allows to make inference about an unobserved process based on the observed one. Notice that the set of the unobserved factors which may affect a country's growth path is potentially large, including institutional setting, political stability, educational system, etc.[9] Using the theoretical model developed above as a guidance, parameter estimates let us infer which effect prevails, complexity or specialization. Notably, this econometric approach provides a classifica-

[9]The Hidden Markov models are a subclass of autoregressive models with Markov regime, for which the conditional distribution of the dependent variable does not depend on its lagged values but only on the regime. For a discussion on the link between Markov-switching models and Hidden Markov models, see Junag and Rabiner (1985) and Rabiner (1989).

tion of our countries' long-run growth rates, in which the growth-regime membership depends on changes in our theoretical growth determinants.

4.1 Econometric strategy

Our econometric strategy is articulated in two steps. In the first step, we estimate the empirical counterpart of equation (4) in order to compute country level measures for Ψ and γ_N. To obtain country-specific estimates for χ_1, χ_2 and λ, we take logs on both sides of equation (4) and regress the growth rate of patents (by residents) on the number of patents, population and number of researchers, for each country i, with $i = 1, \ldots, I$:

$$\left(\frac{\dot{N}}{N}\right)_t = \alpha_0 + \alpha_1 N_t + \alpha_2 L_t + \lambda L_{Nt} + \epsilon_t \tag{13}$$

where $\widehat{\chi}_j = -\widehat{\alpha}_j$, with $j = 1, 2$, and $\epsilon_{it} \sim$ i.i.d $N(0, \sigma_t)$. For each country i, then, we use the OLS estimates $\widehat{\lambda}$, $\widehat{\chi}_1$ and $\widehat{\chi}_2$ to compute the theoretical variable $\gamma_{N,i}$, according to equations (8) and (12).

In the second step, we use the parameters obtained in the first step to estimate the empirical version of equation (9). To exploit the longitudinal nature of our data set and explore the transitions from phases of low growth to phases of high growth and vice versa, we estimate a HMM, which allows to take into account the unobserved heterogeneity. In this perspective, the real GDP growth is interpreted as a result of countries switching between distinct growth regimes (as in Kerekes, 2012). Let $\{\mathbf{\Gamma}_{it}; i = 1, ..., I, t = 1, ..., T\}$ denote sequences of multivariate longitudinal observations for the real GDP growth rate recorded on I countries and T year, where $\mathbf{\Gamma}_{it} = (\gamma_{it1}, ..., \gamma_{itP})^\top \in \mathbb{R}^P$, and let $\{\mathcal{S}_{it}\}$ be a first-order Markov chain defined on the state space $\{1, ..., k, ..., K\}$. A HMM is a stochastic process consisting of two parts: the underlying unobserved process $\{\mathcal{S}_{it}\}$, fulfilling the Markov property, i.e.

$$Pr(\mathcal{S}_{it} = s_{it} | \mathcal{S}_{i1} = si1, \mathcal{S}_{i2} = s_{i2}, \ldots, \mathcal{S}_{it-1} = s_{it-1}) = Pr(\mathcal{S}_{it} = s_{it} | \mathcal{S}_{it-1} = s_{it-1})$$

and the state-dependent observation process $\{\mathbf{\Gamma}_{it}\}$ for which the conditional independence property holds:

$$f\left(\mathbf{\Gamma}_{it}=\gamma_{it}|\mathbf{\Gamma}_{i1}=\gamma_{i1},\ldots,\mathbf{\Gamma}_{it-1}=\gamma_{it-1},\mathcal{S}_{i1}=s_{i1},\ldots,\mathcal{S}_{it}=s_{it}\right)=f\left(\mathbf{\Gamma_{it}}=\gamma_{it}|\mathcal{S}_{it}=s_{it}\right)$$

where $f(\cdot)$ is a generic probability density function (Maruotti and Punzo, 2017). The distribution of $\mathbf{\Gamma}_{it}$ depends only on s_{it}, i.e. $\mathbf{\Gamma}_{it}$ is conditionally independent given the s_{it}. In our baseline specification, we assume that the state-dependent distributions come from a parametric family of continuous or discrete distributions. Thus, the unknown parameters in the HMM involve both the hidden Markov chain and the state-dependent distributions of the random variable $\mathbf{\Gamma}_{it}$. The hidden Markov chain has K states with initial probabilities $\pi_{ik}=Pr(\mathcal{S}_{i1}=k)$, $k=1,\ldots,K$, and transition probabilities

$$\pi_{i,k|j}=Pr(\mathcal{S}_{it}=k|\mathcal{S}_{t-1}=j)$$

with $t=2,\ldots,T$ and $j,k=1,\ldots,K$. For sake of simplicity, we will consider a homogeneous HMM in which common transition and initial probabilities are assumed, i.e. $\pi_{i,k|j}=\pi_{k|j}$ and $\pi_{ik}=\pi_k$, for $i=1,\ldots,I$. The transition probabilities are constant over time and among individuals.[10] The initial probabilities are collected in the K-dimensional vector π, while the time-homogeneous transition probabilities are collected in the $K\times K$ transition matrix $\mathbf{\Pi}$.

The empirical counterpart of the equation for the BGP growth rate of real *per capita* GDP (9) can be written as:

$$E(\gamma_{it}|\mathbf{\Gamma}_{i,t-1},s_{it};\mathbf{m_{it}},\mathbf{n_{it}})=\omega_{0,s_i}+\mathbf{m_i^\top}\omega_{1,s_i}+\mathbf{n_i^\top}\omega_{2,s_i} \tag{14}$$

where, for each country i and time t, $\gamma_{i,t}$ is the GDP growth rate, $\mathbf{m_i^\top}$ is the vector of

[10]This assumption can be easily relaxed to include covariates and/or unit-specific random effects (see, e.g., Maruotti and Rocci, 2012).

the annual products of intermediate sector's markup and γ_N and $\mathbf{n_i^\top}$ is the vector of the annual observations of γ_N. The parameters ω_{2,s_i} and ω_{1,s_i} capture the state-specific unobserved factors that affect GDP growth, through γ_N and its interaction with the intermediate sector's markup, respectively.

4.2 The data

Our final dataset merges information drawn from four different sources. First, we get data on real GDP growth, population growth, and human capital from the Penn World Table database (PWT, hereafter).[11]

Second, in order to construct a measure for the markup in the intermediate sector[12], we use the EUKLEMS database, which provides sector level estimates of markups for a large sample of OECD countries.[13] At the lowest level of aggregation, data are collected for 72 industries according to the European NACE revision 1 classification. As Bucci, Carbonari and Trovato (2019), we assign the following industries to the intermediate sector: *basic metals and fabricated metal; electrical and optical equipment; electricity; gas and water supply; machinery; other non-metallic mineral; rubber and plastics; textile, leather and footwear; transport and storage; transport equipment; wood and cork*. The markup index for the intermediate sector, of country i at time t, is computed as follows:

$$m_{it} = \frac{\textit{Value Added}_{it}}{\textit{Total Labor Costs}_{it} + \textit{Total Capital Costs}_{it}} \tag{15}$$

where all variables are in nominal prices and the base year is 1995.[14] Markup indexes vary across countries (see Table A8) and, on average, have increased along the period

[11]Real GDP is measured in constant 2005 US$. Data source: `http://www.rug.nl/ggdc/productivity/pwt/`. For further details see Timmer et al. (2007).

[12]See Klette (1999), Griffith, Harrison and Macartney (2006) and Nekarda and Ramey (2013).

[13]More information on the EUKLEMS database is available at the web page: `http://www.euklems.net/`.

[14]Results with alternative measures for the intermediate sector's markup are discussed in Paragraph 4.5. Further information on the methodology used in the EUKLEMS database to calculate markups can be found at the link: `https://ec.europa.eu/economy_finance/publications/pages/publication9467_en.pdf`.

under observation (see Figure A3).

Third, the "number of ideas already invented", i.e., a country's stock of knowledge, has been measured through the patent applications by residents, collected by the World Bank.[15] Finally, OECD provides data on the number of researchers at work.[16]

Our final dataset includes 18 OECD countries, over the period 1983-2007. Table A8 presents summary statistics.[17] On average, our data seem to reject the hypothesis of strong scale effect, i.e. $\widehat{\chi}_1 > 0$. Table A9 reports some descriptive statistics on 5-year average real per capita GDP growth rate.

4.3 Results

Our initial state distribution assigns equal probability to all regimes and we let the number of regimes vary between zero (that is a homogenous time dependent process) and five. To determine the number of regimes we use the Bayesian Information Criterion (BIC), which rejects a model without clustering in favor of a model containing four regimes, as in Jerzmanowski (2006) and Kerekes (2012) (see Table A10). On the basis of the 5-years average *per capita* GDP growth rate, we label the regimes as follows: *slow growth*, *steady growth*, *sustained growth* and *miracle growth*.

Table 2: Growth regression, equation (14)

	Constant	$\gamma_{\mathbf{N}}$	$\mathbf{m} \times \gamma_{\mathbf{N}}$	**Cluster standard deviation**	**5-year avg.** ***per capita*** **GDP growth rate (%)**
OLS FE	0.028***	0.023	-0.011		
HMM:					
1–Slow growth	0.821***	0.032***	0.031***	0.712***	1.134
2–Steady growth	2.531***	0.029***	0.007***	0.414***	2.676
3–Sustained growth	3.936***	0.027***	0.006***	0.561***	4.163
4–Miracle growth	5.794***	-0.001***	0.092***	0.997***	6.615

Significance levels: : 10% **: 5% ***: 1%.*

Table 2 presents our main results. The first line presents the OLS fixed effects

[15]Data source: `https://data.worldbank.org/indicator/ip.pat.resd`.

[16]Data source: `https://data.oecd.org/rd/researchers.htm`.

[17]Hungary presents an extremely high value for χ_2 (160.182). When we exclude Hungary from the sample, the sample mean becomes -2.594 while the standard deviation declines to 2.496. Our estimates are robust to the exclusion of this outlier.

estimates of the growth equation while the rest of the table reports the regime-specific parameter estimates; the last column of the table provides the implied long-run growth rates of each regime. The *slow growth* regime is the cluster with the lowest growth rate of GDP (1.134%, standard deviation of 0.712). This regime captures the long lasting stagnation of the Japanese economy, but also the severe downturns in economic activity experienced in France over the period 1998-2006 (see Cette, Fernald and Mojon, 2016), Spain at the end of the 80s (see Blanchard and Jimeno, 1995) and Sweden in the mid-90s as a consequence of the second OPEC crisis (see Jonung and Hagberg, 2005). The *steady growth* regime is characterized by a moderate growth rate of GDP (2.676%) and variability (with a standard deviation of 0.414, this cluster presents the lowest level of variability). This regime captures, for instance, the growth experience of countries like Germany, Australia, Belgium, Canada, Denmark and US. Interestingly, these countries spent most of the time within this regime: this means that, despite cyclical fluctuations, no persistent changes has been found – along the period under observation – in their growth trajectories. The *sustained growth* regime is characterized by higher long-run growth rate (4.163%) and variability (0.561): Austria (continuously from 1989 to 2000) and Spain (continuously from 1993 to 2002) are are the most frequent countries within this regime. Finally, the *miracle growth* regime features the highest long-run growth rate (6.615%) and standard deviation (0.997). This regime takes account of the spectacular growth performance of the Irish economy (which continuously stays within this cluster, from 1990 to 2005) but also of the growth successes of Japan (from 1988-1993), Portugal (in the mid-90s) and Spain (in the first decade of 2000s). Table 3 presents our classification.

The annual growth rate of population is a key ingredient of our theory. In our data, its impact on *per capita* GDP growth is found to be generally positive. In the OLS fixed effects model, $d\gamma/dn$=0.115 (p-value=0.081), while in the HMM we obtain: $d\gamma/dn$=0.389 (p-value=0.000) in the *slow growth* regime, $d\gamma/dn=-0.269$ (p-value=0.000) in the *steady growth* regime, $d\gamma/dn$=0.485 (p-value=0.000) in the *sus-*

Table 3: Classification

Country	**Regime 1** *Slow growth*	**Regime 2** *Steady growth*	**Regime 3** *Sustained growth*	**Regime 4** *Miracle growth*
Australia	4	14	4	0
Austria	4	6	12	0
Belgium	3	14	4	0
Canada	7	12	0	0
Denmark	7	12	3	0
France	12	7	3	0
Germany	1	16	5	0
Hungary	2	6	5	0
Ireland	0	0	4	16
Italy	7	8	7	0
Japan	10	3	2	6
Poland	1	6	0	0
Portugal	3	4	8	4
Slovenia	0	7	2	0
Spain	2	6	10	4
Sweden	5	8	9	0
United Kingdom	9	8	5	0
United States	5	12	0	0

tained growth regime and $d\gamma/dn$=0.083 (p-value=0.000) in the *miracle growth* regime.

We pointed out in Section 3 that, depending on the sign of Φ, the losses due to complexity may be larger or smaller than the gains due to specialization.[18] We also showed that, depending on the sign and magnitude of Φ, innovation and long-run growth may exhibit a weak or even negative correlation, in line with the suggestive evidence provided in Section 1. Notice that the theoretical growth equation (9) and the empirical equation (14) differ because of the presence in the latter of a random intercept, to account for omitted covariates or country-specific heterogeneity, which are not captured by the observed covariates. Thus, in order to obtain an estimate of Φ more directly linkable to the theoretical model, we use the estimates of (14) with $\omega_{0,s_i} = 0$ to get the regime-specific $\widehat{\Phi}_{s_i} = \widehat{\omega}_{1,s_i}\overline{m}_{s_i} + \widehat{\omega}_{2,s_i}$ with $i = 1, \ldots, 4$, where $\overline{m}_{s_i}$ is the regime-mean of the intermediate sector's markup.[19] We find that $\widehat{\Phi}$ is positively correlated with the long-run GDP growth rate, being 0.072 (p-value=0.000) in the *slow growth* regime, 0.090 (p-value=0.000) in the *steady growth* regime, 0.278 (p-

[18] See the discussion at the end of Paragraph 3.2.

[19] To consider the model without intercept in a logitudinal setting, we estimate ω_{1,s_i} and ω_{2,s_i} by Maximum Likelihood.

value=0.000) in the *sustained growth* regime and 0.495 (p-value=0.000) in the *miracle growth* regime. Our theory predicts that a higher Φ implies a stronger positive impact of the proliferation of "new ideas" on the long-run *per capita* GDP dynamics. The switch between different growth regimes, therefore, implies at the same time the transition from phases in which complexity hampers more the GDP growth (lower Φ) to phases in which complexity hampers it less (higher Φ), and vice-versa.[20]

Table 4: Transition matrix

From Regime \ *To Regime*	*1–Slow growth*	*2–Steady growth*	*3–Sustained growth*	*4–Miracle growth*
1–Slow growth	0.488	0.402	0.085	0.024
2–Steady growth	0.245	0.565	0.163	0.027
3–Sustained growth	0.060	0.386	0.530	0.024
4–Miracle growth	0.000	0.000	0.267	0.733

Note: the entry in row j, column i should be interpreted as $p_{ij} = P(st = j|s_{t-1} = i)$.

Table 4 presents the cluster-specific transition probability matrix, which indicates the probability of moving from the column regime to the row regime. The highest persistency is found in the *miracle growth*: when a country (namely, Ireland, Japan, Portugal and Spain) spends a year in this regime, it continues to growth at a so terrific rate in the subsequent year with a probability equal to 0.733; otherwise it moves into the *sustained growth* regime. The *steady growth* and the *sustained growth* regimes also show a high persistency: when a country finds itself in one of the two regimes, the probabilities of remaining there in the subsequent year are 0.565 and 0.530, respectively. At the same time, it is more likely that a country in the *sustained growth* regime slowdowns, moving towards the *steady growth* one (0.386), rather than the opposite occurs (0.163). Finally, a country in the *slow growth* regime will improve its conditions, moving into the *steady growth* regime with a probability of 0.402, the probability of moving into regimes characterized by faster growth being 0.109.

For a better comprehension of our results, Table 5 reports the conditional means for our theoretical growth determinants, whereas Table 6 reports conditional means for

[20]With exclusive reference to the countries in Figures 1 and 2, we find a $\widehat{\Phi}$=0.186 (p-value=0.000).

a list of other structural factors which have been not explicitly taken into account in our theoretical framework. A rich relation emerges between growth and the theoretical growth determinants across the four growth regimes. Not surprisingly, since we are dealing with a sample of OECD countries, we do not observe significant differences in growth fundamentals across clusters. According to our theory, a higher growth rate of patents is found associated with a faster long-run GDP growth, which is also positively correlated with the saving rate and gross capital formation. Interestingly, five non-monotonic relationships with GDP growth rate emerge: the annual population growth rate shows an U-shaped effect on it while the effects of intermediate sector's markup and also λ, χ_1 and χ_2 present an inverted U-shaped pattern. Thus, our evidence suggests that the duplication, difficulty and dilution effects, defined in equation (4), also relate non-linearly with long-run growth. These non-linearities may explain why the underlying positive relationship between patents or researchers and the long-run GDP growth is not clearly found in the aggregate data, although it is predicted by the theory.

Table 5: Key growth-affecting factors, conditional means

Regime	**m**	$\widehat{\lambda}$	$\widehat{\chi_1}$	$\widehat{\chi_2}$	$\dot{\mathbf{N}}/\mathbf{N}$	**n**
1–Slow growth	1.061	0.011	0.250	-0.520	6.073	0.531
2–Steady growth	1.193	0.011	0.273	4.413	10.442	0.517
3–Sustained growth	1.127	0.184	0.264	-2.917	11.868	0.398
4–Miracle growth	1.051	-0.157	0.248	-4.193	12.496	0.659

Table 6: Other growth-affecting factors, conditional means

Regime	**saving rate**	**investment share**	**public spending share**
1–Slow growth	0.424	0.253	0.170
2–Steady growth	0.423	0.259	0.168
3–Sustained growth	0.423	0.277	0.169
4–Miracle growth	0.484	0.280	0.143

4.4 Other growth fundamentals

Finally, we estimate several specifications of the Multinomial Logit Model (MNL) to assess the role of the other potential long-run growth fundamentals in affecting a country cluster's membership. In this exercise, we take the *slow growth* regime as reference. Therefore, the multinomial logistic regression evaluates the relative probability of being in one of the remaining growth regimes against the reference, using a linear combination of predictors. The obtained MLE-estimated coefficients represent the effects of every predictor variable in the log-odds of being in any other regime versus the reference regime. As predictor variable we employ the human capital index (`hc`) provided by PWT. Results are reported in Table 7. *Ceteris paribus*, a unit increase in the human capital index increases the probability of being in the *steady growth* regime, relative to the *slow growth* regime, by a multiplicative factor of exp(0.191)=1.210, i.e. increasing it by 21% (p-value=0.000). This confirms that, even for high income countries, human capital is one of the underlying factors that lead to the transition from stagnation to growth. The increase in educational level measured by the human capital index, however, does not explain the take-off toward regimes characterized by higher growth rates, being this ultimately related to R&D intensity and specialization.

Table 7: MNL Model for regime membership

From *slow growth* to...	**Coef.**	**Std.Err.**
... steady growth		
hc index	0.191***	0.014
constant	0.132	1.658
... sustained growth		
hc index	-1.741***	0.111
constant	5.424***	1.661
... miracle growth		
hc index	-2.879***	0.143
constant	7.647***	1.961

4.5 Robustness checks

In this paragraph, we briefly discuss how estimates and classification behave in response to changes in the econometric specification and/or in the way we measure some explanatory variables.[21]

Real GDP growth rate. First, we examine whether our results are sensitive to alternative measures of GDP growth rate. We run our regressions using as dependent variable, respectively, (i) the rate of growth of per worker GDP, using the number of persons engaged (`emp`), and (ii) the rate of growth of per hours worked GDP, using the average annual hours worked (`avh`) (source: PWT, see Figure A4). In this way, besides a slight change in the composition of the cluster, we observe that the estimates appear less accurate.

$\widehat{\boldsymbol{\Psi}}$. We check the robustness of our results using an alternative estimate of Ψ. For each country i, with $i = 1, \ldots, I$, we regress the growth rate of patents (by residents) on the demographic growth rate, according to (8):

$$\left(\frac{\dot{N}}{N}\right)_t = \Psi n_t + \epsilon_t \tag{16}$$

with $\epsilon_{it} \sim$ i.i.d $N(0, \sigma_t)$. For each country, then, we use the OLS estimates $\widehat{\Psi}$ to compute the theoretical variable $\gamma_{N,i}$. Despite some modifications occur in the composition of the clusters, we still identify four growth regime. The regressions, however, result in less significant estimates and lower explanatory power.

$\widehat{\chi}_{\mathbf{1}}$ **and** $\widehat{\chi}_{\mathbf{2}}$. The HMM rests on the fact that the growth regime transition probabilities are time dependent. A theoretical explanation for having (at least some) time dependent parameter in our model, can be found in the possible interplay between N and L_N. To take this into account, we modify equation (4) as follows:

$$\dot{N}_t = \frac{1}{X} \cdot N_t^{1-\chi_1(N_t)} \cdot L_t^{-\chi_2(N_t)} \cdot L_{Nt}^{\lambda}, \tag{5'}$$

The exponents $\chi_i(N)$, $i = 1, 2$, are now intended as (factor-specific) complexity indices pertaining to R&D activity and depend positively on N, in line with Sequeira, Gil and

[21] As our results survive various robustness checks, for the sake of brevity we do not present and discuss in detail all the parameter estimates, which are available upon request.

Afonso (2018). As the number of varieties of specialized (intermediate) goods rises, the economy becomes more complex because there is an increase in the diversity (and maybe redundancy) of the different components (i.e., intermediate goods) that need to be assembled for producing the final output. This, in turn has two effects. Concerning $\chi_1(N_t)$, we postulate that a proliferation of N_t amplifies the difficulty with which new ideas are discovered starting from the oldest ones. Similarly, as for $\chi_2(N_t)$, we maintain that the same increase in N_t strengthens the effect by which a rise in population size ultimately leads to a dilution of the total amount of R&D expenditures over a larger number of (more) dispersed research projects. However, we also assume that the indices will eventually level off, i.e., $\chi_i(N_t) \to \chi_i$, $i = 1, 2$, despite the continuous increase in the number of varieties, reflecting the fact that part of the modern innovations (leading to new varieties of goods) have a stabilizing role in the complexity of the economies.[22] The empirical counterpart of equation (5') is given by the following reduced form equation:

$$\left(\frac{\dot{N}}{N}\right)_t = \alpha_0 + \alpha_1 N_t + \alpha_2 N_t L_t + \lambda L_{Nt} + \epsilon_t \tag{17}$$

where $\widehat{\chi}_j = -\widehat{\alpha}_j$, with $j = 1, 2$, and $\epsilon_{it} \sim$ i.i.d $N(0, \sigma_t)$.The results of the HMM model obtained using these alternative $\widehat{\chi}_1$, $\widehat{\chi}_2$ and $\widehat{\lambda}$ are in line with those presented above. We still identify four growth regimes: *slow growth* (with 5-years average growth rate of 0.994%), *steady growth* (2.654%), *sustained growth* (3.826%) and *miracle growth* (6.371%). Clusters' standard deviations follow the same pattern of those in the baseline model and only slight modifications occurs in clusters' composition. Estimates, however, are less accurate (see Table A11).

m. We carry out several robustness checks for the markup. We compute it, using

[22]Sequeira, Gil and Afonso (2019, p. 107) point out that, at least in modern economies, "*some inventions have reduced or attenuated the effect of complexity either directly or indirectly. Computers have allowed calculation of things and analysis of data and models in ways impossible before, while, as an indirect effect, one could argue that the development of machinery that replaces humans (e.g., earth-moving equipment) have allowed more humans to spend their time managing complexity and its effects*". Their empirical results indeed suggest a stabilization of complexity over the long-run.

an alternative base year, the 1985, and we do not find any significant change in our results. To tackle the measurement issue, we follow Roeger (1995), running the following regression:

$$SR_t - SRP_t = \left(1 - \frac{1}{\tilde{m}}\right) [(\Delta p_t + \Delta Q_t - u_t) + (\Delta r_t + \Delta K_t - v_t)] \tag{18}$$

where: $(SR_t - SRP_t)$ is the difference between the Solow residual and the price-based Solow residual, $\tilde{m}$ is the intermediate sector's markup, $(\Delta p_t + \Delta Q_t)$ is the nominal output growth, $(\Delta r_t + \Delta K_t)$ is the growth of capital cost. In equation (18), both capital costs and nominal output are measured with error, with $v_t \sim i.i.d.(0, \sigma_v)$ and $u_t \sim i.i.d.(0, \sigma_u)$.[23] Estimates for $\tilde{m}$ are presented in Table A12. However, when we estimate our HMM, the lack of time variability of $\tilde{m}$ implies less accurate estimates.

Human capital. We estimate a version of the HMM in which the growth rate of the human capital (proxied by the human capital index provided by PWT, `hc`) replaces the population growth rate (see Bucci, 2015). In the OLS fixed effects model we get that $d\gamma/d\tilde{n}$=0.089 (p-value=0.001), where $\tilde{n}$ denotes the rate of change of `hc`, while no significative changes are observed in the HHM.

5 Concluding remarks

In this paper, we advance an explanation for a cross-country evidence which is inconsistent with most endogenous growth models: the coexistence of increasing trends in aggregate research effort and the no corresponding increases in long-run *per capita* income growth. Building on Jones (1995), we develop a model in which long-run growth is determined by the interplay between "complexity parameters", "technological parameters"and the (exogenous) population growth rate. In this setting, the coexistence of slow *per capita* income growth and the proliferation of "new ideas"is shown to be

[23]See Christopoulou and Vermeulen (2012) for an application to the Euro area countries and the US, over the period 1981-2004.

the equilibrium response to greater complexity.

Using the theoretical restrictions on the BGP growth equation, for a sample of OECD countries, we estimate a Hidden Markow Model, which allows for an endogenous classification mechanism. We find four different regime for the long-run *per capita* real GDP growth. Each country's growth pattern is the result of transitions between distinct growth regimes. The transitions are determined by regime-specific transition probabilities. In light of our classification, we further establish several facts about the transition between different growth regimes. We find that growth accelerations, in high income countries, are strongly associated with patenting activity, i.e. the annual growth rate of patent applications (by residents); growth failures meanwhile are characterized by a weaker degree of *specialization*, i.e. a lower average estimated value for Φ. In this situation, in which complexity in production is more harmful for long-run growth, our theory suggests that increases in the population growth rate allow to achieve faster economic growth. An alternative way-out, which our model can easily be extended to account for, is to increase the investment in education: using a Multinomial Logit Model, we find that the level of human capital – here proxied by the human capital index provided by PWT – positively affects the transition from a regime with modest growth towards a regime of faster growth.

Our results regarding the interplay between specialization and complexity yield important policy implications inasmuch as they show that increases in the population growth rate can help countries (namely those stuck in a *slow growth* regime) achieve higher long-run economic growth rates. A higher population growth rate can be achieved in multiple ways, ranging from policies that improve health and fertility of people to pro-migration ones. By the same token, if we interpret the labor input in the R&D sector in our model as human-capital factor (similarly to, e.g., Romer, 1990), then policies aimed at increasing the growth rate of the labor force with high human capital, in particular of the kind adequate for science, technology, engineering, and mathematics (STEM) occupations, should also help countries attain higher long-run economic growth. In this sense, the whole of such policies may be viewed within our framework as fully complementary to the well-known innovation policies already prescribed by R&D-based growth models. Last but not least, our results also point to the relevance of policies that promote general investment in education (say, to increase overall school attainment) as a higher level of human capital seems to facilitate the transition to higher growth regimes.

References

[1] Aghion, Philippe and Peter Howitt (1998), *Endogenous Growth Theory*, Cambridge, MA: The MIT Press.

[2] Ang, J. B. and J. B. Madsen, (2015) What Drives Ideas Production across the World? *Macroeconomic Dynamics*, 19(1), 79-115.

[3] Backus, D., P. Kehoe, and T. Kehoe (1992). In search of scale effects in trade and growth. *Journal of Economic Theory*, 57, 377-409.

[4] Bassanini, A., S. Scarpetta, and I. Visco (2000). Knowledge, technology and economic growth: Recent evidence from OECD countries. *OECD Economics Department Working Papers* 259, 1-38.

[5] Baum, Leonard. E. and J.A. Eagon (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73(3), 360-363.

[6] Baum, Leonard. E. and and Ted Petrie (1966). Statistical inference for probabilistic functions of finite-state Markov chains. *Annals of Mathematical Statistics*, 37(6), 1554-1563.

[7] Benassy, Jean-Pascal (1996a) Taste for Variety and Optimum Production Patterns in Monopolistic Competition. *Economics Letters* 52(1), 41-47.

[8] Benassy, Jean-Pascal (1996b) Monopolistic Competition, Increasing Returns to Specialization and Output Persistence. *Economics Letters* 52(2), 187-191.

[9] Benassy, Jean-Pascal (1998) Is there always too little research in endogenous growth with expanding product variety? *European Economic Review* 42(1), 61-69.

[10] Blanchard, Olivier and Juan F. Jimeno (1995) Structural Unemployment: Spain versus Portugal. *The American Economic Review, Papers & Proceedings* 85(2), pp. 212-218.

[11] Bucci, Alberto (2013) Returns to Specialization, Competition, Population, and Growth. *Journal of Economic Dynamics and Control* 37(10), 2023-2040.

[12] Bucci, Alberto (2015) Product Proliferation, Population, and Economic Growth. *Journal of Human Capital* 9(2), 170-197.

[13] Bucci, Alberto, Lorenzo Carbonari and Giovanni Trovato (2019) Variety, Competition, and Population in Economic Growth: Theory and Empirics. *Macroeconomic Dynamics*, forthcoming.

[14] Cette, Gilbert, John Fernald and Benoît Mojon (2016) The pre-Great Recession slowdown in productivity. *European Economic Review* 88, pp. 3-20.

[15] Dinopoulos, Elias, and Peter Thompson (1998) Schumpeterian Growth without Scale Effects. *Journal of Economic Growth*, 3, 313-335.

[16] Dinopoulos, Elias and Peter Thompson (1999) Scale effects in Schumpeterian models of economic growth. *Journal of Evolutionary Economics* 9(2), 57-185.

[17] Dinopoulos, Elias and Peter Thompson (2000) Endogenous growth in a cross-section of countries. *Journal of International Economics*, 51(2), 335-362.

[18] Ethier, Wilfred J. (1982) National and International Returns to Scale in the Modern Theory of International Trade. *American Economic Review* 72(3), 389-405.

[19] Ferrarini, Benno and Pasquale Scaramozzino (2016) Complexity, Specialization and Growth. *Structural Change and Economic Dynamics* 37, 52-61.

[20] Gil, P.M., P. Brito, and O. Afonso (2013). Growth and firm dynamics with horizontal and vertical R&D. *Macroeconomic Dynamics*, 17(7), 1438-1466.

[21] Griffith, Rachel, Rupert Harrison and Gareth Macartney (2006) Product Market Reforms, Labour Market Institutions and Unemployment. *Economic Journal* 117, 142-166.

[22] Grossman, Gene and Elhanan Helpman (1991) *Innovation and Growth in the Global Economy*, MA: MIT Press.

[23] Hausmann, Ricardo and Bailey Klinger (2006) Structural Transformation and Patterns of Comparative Advantage in the Product Space. *KSG Working Paper No. RWP06-041*, Harvard University: Center for International Development.

[24] Howitt, Peter (1999) Steady Endogenous Growth with Population and R&D Inputs Growing, *Journal of Political Economy*, 107, 715-730.

[25] Jerzmanowski, Michay (2006) Empirics of hills, plateaus, mountains and plains: A Markov-switching approach to growth. *Journal of Development Economics* 81, 357- 385.

[26] Jones, Charles I. (1995) R&D-based Models of Economic Growth. *Journal of Political Economy* 103, 759-84.

[27] Jones, Charles I. (1999) Growth: With or Without Scale Effects? *American Economic Review*89, 139-44.

[28] Kerekes, Monika (2012) Growth Sustaineds and failures in a Markov switching classification model of growth. *Journal of Development Economcs* 98, 167-177.

[29] Klette, Tor J. (1999) Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data. *Journal of Industrial Economics* 47(4),451-476.

[30] Kortum, Sam (1997) Research, Patenting, and Technological Change, *Econometrica*, 65, 1389-1419.

[31] Kremer, Michael (1993) The O-Ring Theory of Economic Development. *Quarterly Journal of Economics* 108(3), 551-575.

[32] Laincz, C. A. and Peretto, P. F. (2006) Scale Effects in Endogenous Growth Theory: An Error of Aggregation Not Specification. J*ournal of Economic Growth*, 11, 263-288.

[33] Madsen, J. B. and J. B. Ang (2011) Can Second-Generation Endogenous Growth Models Explain the Productivity Trends in the Asian Miracle Economies? *Review of Economics and Statistics*, 93, 1360-1373.

[34] Madsen, J. B. (2008) Semi-endogenous versus Schumpeterian Growth Models: Testing the Knowledge Production Function using International Data. *Journal of Economic Growth*, 13(1), 1-26

[35] Madsen, J. B., (2010) The Anatomy of Growth in the OECD since 1870. *Journal of Monetary Economics*, 57, 753-767.

[36] Maruotti, Antonello, and Roberto Rocci (2012) A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in Medicine* 31 (9), 871-886.

[37] Maruotti, Antonello, and Antonio Punzo (2017) Model-based time-varying clustering of multivariate longitudinal data with covariates and outliers *Computational Statistics & Data Analysis*, 113, 475-496.

[38] Nekarda, Christopher J., and Valerie A. Ramey (2013) The Cyclical Behavior of the Price-Cost Markup. *NBER Working Paper*, No. 19099.

[39] Peretto, Pietro (1998) Technological Change and Population Growth. *Journal of Economic Growth*, 3, 283-311.

[40] Peretto, Pietro, and Sjak Smulders (2002) Technological Distance, Growth and Scale Effects. *Economic Journal*, 112, 603-624.

[41] Pintea, Mihaela and Peter Thomphson (2007) Technological complexity and economic growth. *Review of Economic Dynamics* 10, 276-293.

[42] Rabiner, Lawrence R. (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77, 257-286.

[43] Roeger, Werner (1995) Can Imperfect Competition Explain the Difference Between Primal and Dual Productivity Measures? Estimates for US Manufacturing. *Journal of Political Economy* 103, 316-330.

[44] Romer, Paul M. (1990) Endogenous Technological Change. *Journal of Political Economy*, 98(5), 71-102.

[45] Sala-i-Martin, Xavier, Gernot Doppelhofer and Ronald I. Miller (2004) Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94(4), 813-835.

[46] Sequeira, Tiago Neves, Pedro Mazeda Gil and Óscar Afonso (2018) Endogenous growth and entropy. *Journal of Economic Behavior & Organization*, 154, 100-120.

[47] Segerstrom, Paul S. (1998) Endogenous Growth Without Scale Effects. *American Economic Review*, 88, 1290-1310.

[48] Timmer, Marcel, Ton van Moergastel, Edwin Stuivenwold, and Gerard Ypma, M. (2007) *EUKLEMS Growth and productivity accounts. Part I: methodology*, The Netherlands: University of Groningen & University of Birmigham.

[49] Young, Alwyn (1998) Growth Without Scale Effects. *Journal of Political Economy*, 106, 41-63.

A Appendix

A.1 Derivation of the theoretical results

The proof of point 1) is straightforward since we assume $n > 0$. The proof of point 2) is also immediate by observing that: $\frac{\partial \gamma_y}{\partial n} = \Psi\Phi$. To prove point 3), notice that, in equation (9), $\gamma_y = 0$ if $n = 0$. Finally, to prove point 4), notice that equation (9) also implies that:

$$\frac{\partial \gamma_y}{\partial m} = \begin{cases} \Psi n\left[\alpha(1-\beta)\right] > 0 & \text{if} \quad (\Psi > 0 \,\text{and}\, \beta < 1) \,\text{or}\, (\Psi < 0 \,\text{and}\, \beta > 1) \\ \Psi n\left[\alpha(1-\beta)\right] = 0 & \text{if} \quad \Psi = 0 \,\text{or}\, \beta = 1 \\ \Psi n\left[\alpha(1-\beta)\right] < 0 & \text{if} \quad (\Psi < 0 \,\text{and}\, \beta < 1) \,\text{or}\, (\Psi > 0 \,\text{and}\, \beta > 1) \end{cases}$$

■

Table A8: Summary statistics

Variable	Description	Obs.	Mean	Std. dev.	Min	Max
γ_y	5-year avg. real *per capita* GDP growth rate	344	3.011	1.695	-0.938	9.045
m	intermediate sector's markup index (1995=1)	344	1.133	0.503	0.719	4.794
n	annual population growth rate	344	0.504	0.488	-0.290	2.205
N	number of patents by residents	344	14,177.910	23,380.950	6.000	12,8152.200
L_N	number of researchers	344	12,745.67	14,257.49	251	61800
$\widehat{\lambda}$	country-by-country OLS estimate for λ	344	0.038	0.845	-0.979	3.865
$\widehat{\chi}_1$	country-by-country OLS estimate for χ_1	344	0.263	0.237	0.061	1.332
$\widehat{\chi}_2$	country-by-country OLS estimate for χ_2	344	0.718	23.148	-9.271	160.182

Table A9: Descriptive statistics on 5-year average real *per capita* GDP growth rate

Country	**Min**	**Mean**	**Max**
Australia	0.942	2.526	3.560
Austria	1.579	3.250	5.114
Belgium	-0.938	2.617	5.633
Canada	-0.228	1.943	3.395
Denmark	1.485	2.590	3.888
France	-0.275	1.987	4.242
Germany	1.358	3.022	5.414
Hungary	1.084	2.949	5.015
Ireland	3.569	6.534	9.045
Italy	-0.338	2.401	4.938
Japan	-0.419	2.968	7.747
Poland	1.590	3.354	6.361
Portugal	0.580	4.015	7.325
Slovenia	2.046	2.719	3.894
Spain	-0.828	3.762	6.349
Sweden	-0.445	2.643	4.410
United Kingdom	1.311	2.767	4.955
United States	1.248	2.310	3.474

Table A10: Information criteria

	2 regimes	**3 regimes**	**4 regimes**	**5 regimes**
Log-likelihood	-566.69	-501.56	-463.24	-439.53
AIC	1155.4	1043.1	988.48	967.07
BIC	1197.6	1119.9	1107.5	1136.1

Table A11: Robustness check – growth regression, using equation (17) to get $\widehat{\chi}_1, \widehat{\chi}_2$ and $\widehat{\lambda}$

	Constant	$\gamma_{\mathbf{N}}$	$\mathbf{m} \times \gamma_{\mathbf{N}}$	**Cluster standard deviation**	**5-year avg.** ***per capita*** **GDP growth rate (%)**
OLS FE	0.031***	1.025***	-0.561***		
HMM:					
1–Slow growth	1.060***	0.925	-0.886	0.741	0.994
2–Steady growth	2.608***	0.256	-0.257	0.493	2.654
3–Sustained growth	4.095***	0.942	-0.454	0.590	3.826
4–Miracle growth	5.651***	1.467**	5.106**	1.091	6.371

Significance levels: : 10% **: 5% ***: 1%.*

Table A12: Estimated intermediate sector's markup

Country	$\tilde{m}$
Australia	1.665
Austria	1.510
Belgium	1.650
Canada	1.755
Denmark	1.572
France	1.569
Germany	1.665
Hungary	1.620
Ireland	1.518
Italy	1.599
Japan	1.540
Poland	1.596
Portugal	1.683
Slovenia	1.906
Spain	1.608
Sweden	1.538
United Kingdom	1.686
United States	1.713
mean	1.635

Note: all parameters are significant at 1%.

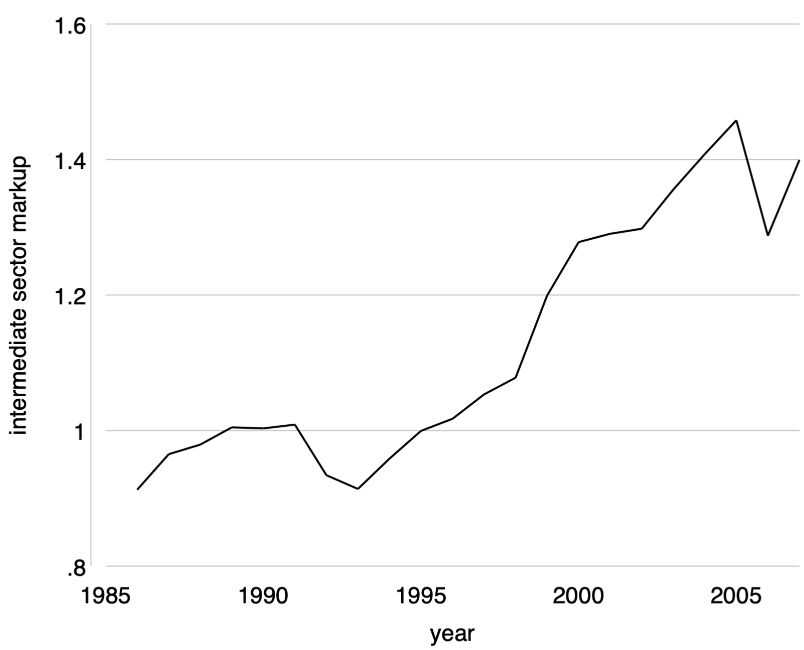

Figure A3: Markup index for the intermediate sector (whole sample)

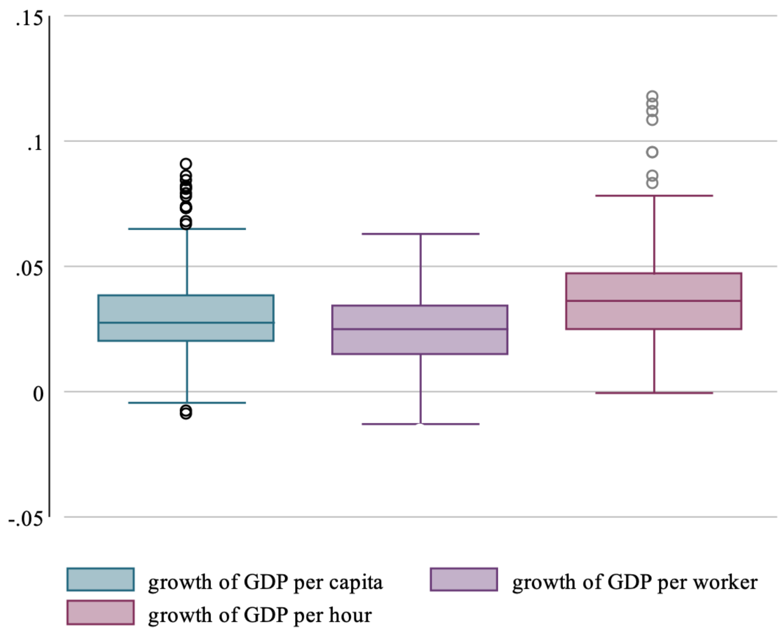

Figure A4: Different measures of GDP growth rates (our sample)