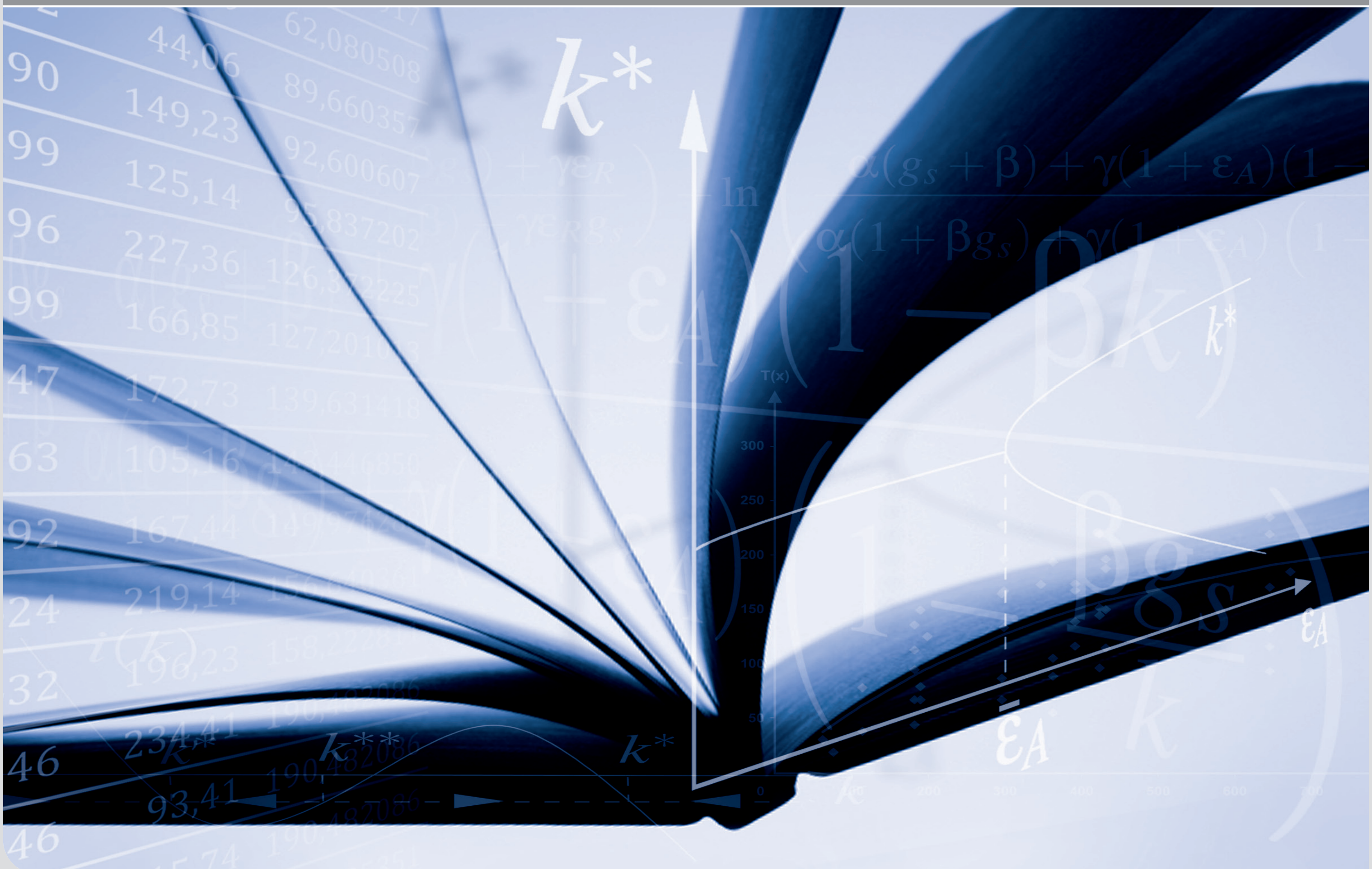
Karlsruher Institut für Technologie

Schumpeterian growth with technological interdependence: An application to US states

by Tim Deeken

No. 75 | NOVEMBER 2015

WORKING PAPER SERIES IN ECONOMICS



KIT – University of the State of Baden-Wuerttemberg and
National Laboratory of the Helmholtz Association

econpapers.wiwi.kit.edu

Impressum

Karlsruher Institut für Technologie (KIT)
Fakultät für Wirtschaftswissenschaften
Institut für Volkswirtschaftslehre (ECON)

Schlossbezirk 12
76131 Karlsruhe

KIT – Universität des Landes Baden-Württemberg und
nationales Forschungszentrum in der Helmholtz-Gemeinschaft

Working Paper Series in Economics
No. 75, November 2015

ISSN 2190-9806

econpapers.wiwi.kit.edu

Schumpeterian Growth with Technological Interdependence: An Application to US States

Tim Deeken*

November 9, 2015

Abstract

In this paper, the Schumpeterian growth model developed by Ertur and Koch (2011) that includes spatial interactions between units of observation working via R&D spillovers is presented in detail. The implications of this model and three additional growth models with and without spatial interaction that are nested within this framework are tested for the US states econometrically. It is found that investments in R&D have a positive impact on steady-state income per worker in the Schumpeterian growth model without complex interaction between states, but this effect is absent in the model proposed by Ertur and Koch (2011), even though the estimate for the coefficient measuring interconnectedness between regions is positive and significant. This latter result is robust to alternative specifications of the interaction matrix.

Keywords: R&D spillovers; Schumpeterian growth; spatial econometrics

JEL: O18, O47, R11, C31

*Karlsruhe Institute of Technology, Institute for Economics, Schlossbezirk 14, 76 131 Karlsruhe, Germany. e-mail: tim.deeken@kit.edu, phone: +49 (0) 721 608-43073, fax: +49 (0) 721 608-45255

1 Introduction

Interaction between countries or regions occurs in many forms. One particular dimension of this interaction concerns the diffusion of knowledge or knowledge spillovers. Given that knowledge is a key factor in economic development, this implies that the level of development, measured by, for example, income per capita, in one region depends on characteristics in the regions it interacts with.

The presence of these interdependent relationships motivates the need to incorporate these also in theoretic models. One example in the recent economic literature that takes into account interdependence between countries is the exogenous growth model developed in Ertur and Koch (2007). In this paper, the transition to an endogenous growth model will be made by presenting a model by the same authors (Ertur and Koch, 2011), which builds heavily upon the contributions by Aghion and Howitt (1998) and Howitt (2000). The novelty of the model by Ertur and Koch (2011) is that it incorporates complex spatial interactions, modeled via technological interdependence between regions, in the context of an endogenous growth model, in which profit-driven investment in research and development (R&D) determines the rate of technological progress. In particular, the authors develop an integrated theoretical and empirical framework that nests a series of growth models.

This paper fills a gap in the literature as, to the best of my knowledge, the model has not yet been investigated empirically for the US states. The shift in focus from a cross-country to a cross-regional setting is important for the following reasons. First, the United States is the global leader in investments in R&D. In 2011, R&D investments in the United States accounted for approximately 30% of the global total, far ahead of the next-ranked countries China, Japan, and Germany with shares of 15%, 10%, and 7%, respectively. The dominance was even higher in 2001, when the United States' global share was 37% (all figures are from National Science Board (2014, 4-17)). The second reason for choosing US states as the units of analysis addresses interdependence between these units. As Keller (2002) points out, the strength of technological knowledge spillovers declines with the geographic distance between the originating and the receiving country, implying that diffusion of technology is not a frictionless process. Geographic distance in this situation captures, for instance, socio-economic differences, but also those in institutions between countries (Ertur and Koch, 2007, 1036), which have been highlighted in the growth literature as a fundamental determinant of cross-country income differences.[1] The advantage of studying diffusion of technology within a single country is the common

[1] As a starting point, consider the seminal contribution by Acemoglu et al. (2001).

institutional setting,[2] which possibly reduces part of the frictions. The third reason for choosing the United States relates back to the first. Eaton and Kortum (1996) find that for the OECD countries the amount of a country's growth in productivity that depends on research efforts in the United States is larger than 50%, which points to substantial spillovers from the United States. In addition, Eaton and Kortum (1999, 558) estimate that in the past 60% of the United States' productivity growth originated from research conducted domestically. This figure is in stark contrast to the corresponding values for Japan or Germany, with figures of 16% or 35%, respectively, and it raises the question, if significant spillovers also exist between US states or only between the Unites States and other countries. The cited figure of 60% is silent about any spillovers between US states.

Indicative evidence for the potential existence of these spillovers is provided by the map in Figure 1, which shows the average R&D investment rate (or R&D intensity) over the period 1997-2007 in the 48 continental US states plus the District of Columbia.[3] States

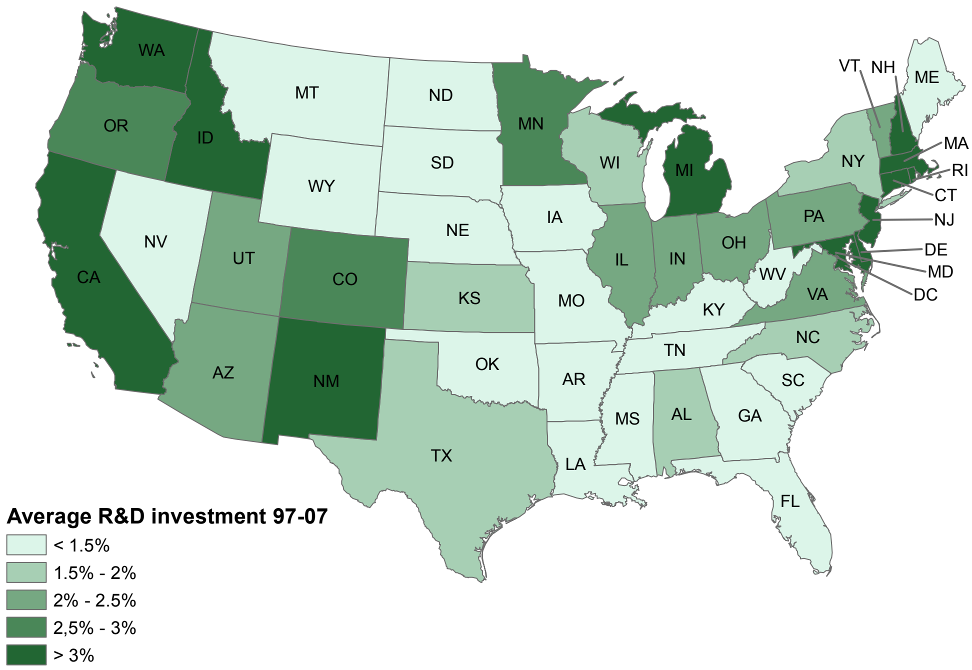


Figure 1: Average R&D Investment Rate for the 48 Continental US States plus Washington, D.C. over the period 1997-2007 (Data: OECD, 2015).

with average R&D investment rates above 2% can be found predominantly on the western seaboard and in the south-west (with the exception of Nevada) as well as in the north-east and the region around the Great Lakes (the notable exception in these regions is Maine).

[2]However, US states have considerable autonomy in the United States' federalist system. For a comparison with the German system on this aspect, see Halberstam and Hills, Jr. (2001).

[3]A correspondence between state abbreviations and names is provided in Table E.6 in Appendix E.

In regions where states with high R&D intensities abound, the potential for spillovers and a subsequent impact on output is high.

Figure 2 illustrates the data on the R&D intensities in a different way. It shows a Moran scatterplot of the average R&D intensity (in standardized form) on the horizontal axis, and the vertical axis measures the standardized value of the spatial lag of this variable. The value of the spatial lag for a given state comprises the average of the R&D intensities of this states' direct neighbors.

Each point in the scatterplot corresponds to a single state so that states in the upper right quadrant have average R&D intensities above the mean and are also surrounded by states for which the same holds. The reverse holds for states in the lower left quadrant, whereas in the upper left quadrant states can be found whose own average R&D investment rate is below the mean, but who are neighbors to states with above-average R&D intensities. The dot labeled "NM" in the figure denotes New Mexico, which has the highest R&D intensity in the sample.[4] However, its neighboring states fall below the average.

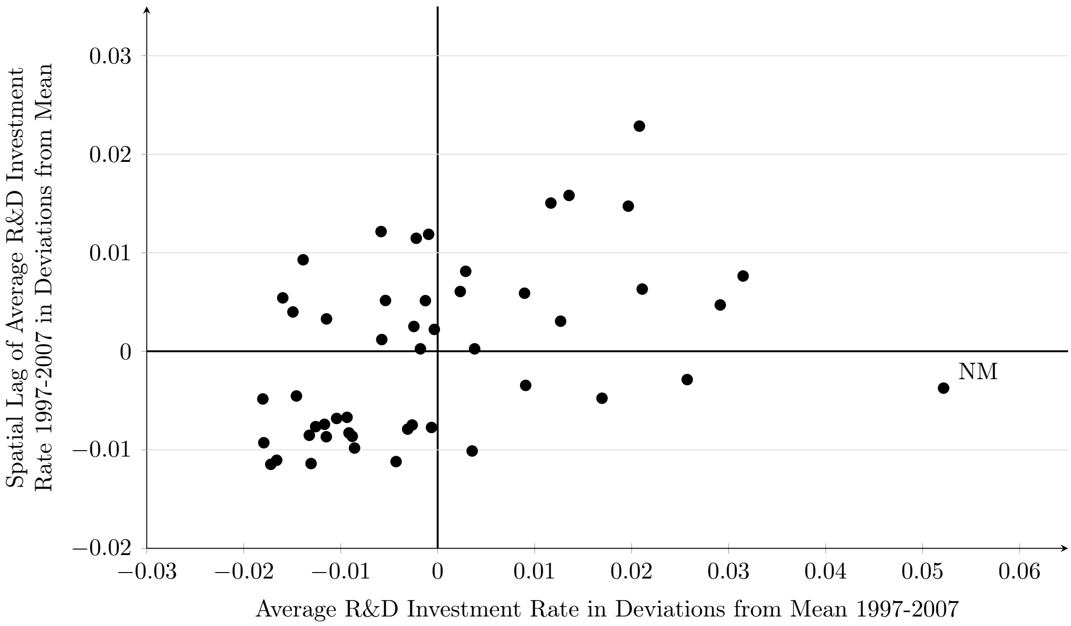


Figure 2: Moran Scatterplot of the Average R&D Investment Rate for the 48 Continental US States plus Washington, D.C. over the period 1997-2007 (Data from (OECD, 2015)).

Note: The variables are in the form of deviations from the mean so that the value 0 on the abscissa is equivalent to the mean value of 2.2%.

[4]The high value for New Mexico can be explained by the presence of Los Alamos National Laboratory and Sandia National Laboratories, which are federally funded research and development centers. Compare the information by the National Science Foundation available under `http://www.nsf.gov/statistics/infbrief/nsf02322/` (accessed: 9 August, 2015).

The paper is organized as follows: Section 2 introduces the basic structure of the multi-region Schumpeterian before Section 3 specifies the nature of technological interdependence between regions and derives the equation for the income per worker in steady state. In Section 4, the focus is on the empirical specification of the model and the estimation strategy. The data for the empirical analysis is presented in detail in Section 5, which also discusses the estimation results. Finally, Section 6 concludes.

2 Multi-Region Schumpeterian Growth Model without Technological Interdependence

This section describes the multi-region Schumpeterian growth model in Ertur and Koch (2011), which builds upon work by Aghion and Howitt (1998, Chapter 3 and 12.2) and Howitt (2000). The expression “multi region” that is attached to this setup might be a slight misnomer though, as each region is assumed to develop independently from the other regions so that the term “single-region model” would be more appropriate for this section. However, to make the transition to the multi-region model in Section 3 easier, already here a single region in the economy will be indexed. Section 2.1 describes the production side of the region’s final good sector and Section 2.2 illustrates its intermediate goods sector, before Section 2.3 clarifies the connections in the research and development (R&D) sector.

2.1 Final Good Sector

The economy under consideration consists of $i = 1, \ldots, N$ regions. A single final good is produced in each region with labor and a continuum of intermediate goods (or varieties) as input factors. The final good sector operates under perfect competition, and the good is produced via the following production function, illustrated here for region i,

$$Y_i(t) = Q_i(t)^{\alpha-1} \int_0^{Q_i(t)} A_i(v,t) x_i(v,t)^{\alpha} L_i(t)^{1-\alpha} \, dv, \tag{1}$$

where $Y_i(t)$ is output in region i at time t. This output, besides its use as a consumption good, also functions as a capital good in the production of intermediates and as an input into research and development activities. The variable $x_i(v,t)$ measures the flow of intermediate good v used in the production of the final good, and $Q_i(t)$ indicates how many different intermediate goods exist in region i at time t. The continuum of intermediates is therefore measured on the interval $v \in [0, Q_i(t)]$. $A_i(v,t)$ is a productivity parameter, which reflects the quality of intermediate product v and thus increases with successive vintages of the good. Finally, $L_i(t) = L_i(0)e^{n_i t}$ is the flow of labor, and $n_i > 0$ is the constant growth rate of labor.[5] It is assumed that the population and labor force size coincide and that labor is supplied inelastically.

Following Acemoglu (2009, 435 and 461), the demand for intermediate good v can be calculated by maximizing the instantaneous profits of a representative final goods producer at time t.[6] The problem is

$$\begin{aligned} \max_{x_i(v,t)} \Pi_i(v,t) = Q_i(t)^{\alpha-1} &\int_0^{Q_i(t)} A_i(v,t)x_i(v,t)^{\alpha} L_i(t)^{1-\alpha}\, dv \\ &- \int_0^{Q_i(t)} p_i(v,t)x_i(v,t)\, dv - w_i(t)L_i(t). \end{aligned} \tag{2}$$

Applying the rule for differentiating under the integral sign, and solving the necessary condition for $p_i(v,t)$ leads to the inverse demand schedule for variety $v \in [0, Q_i(t)]$[7]

$$p_i(v,t) = \alpha A_i(v,t) l_i(t)^{1-\alpha} x_i(v,t)^{-(1-\alpha)}. \tag{3}$$

Here, $l_i(t) \equiv \frac{L_i(t)}{Q_i(t)}$ denotes the number of workers per variety. With the help of results developed in Section 2.2.1, it can be shown that the production function in intensive form is given by[8]

$$\hat{y}_i(t) = \hat{k}_i(t)^{\alpha} \tag{4}$$

where $\hat{y}_i(t) \equiv \frac{Y_i(t)}{A_i(t)L_i(t)}$ is the output per effective worker, and $\hat{k}_i(v,t)$ is capital per effective worker.

Concerning the production function in Equation (1), it is important to note that the integral is multiplied by the factor $Q_i(t)^{\alpha-1}$. The factor is introduced in order to avoid that producers of the final good become increasingly more productive simply due to the

[5] The restriction that the labor growth rate is positive is not actually spelled out explicitly in Ertur and Koch (2011) though. However, since labor is an essential input in the production of the final good (for the proof, see, for example Barro and Sala-i-Martin (2004, 77-78)), the positive growth rate can be inferred. This assumption is maybe not as innocuous as it seems, in particular when it comes to testing the model's implications empirically. See also Footnote 46 in this context.

[6] Again this is not explicitly spelled out in Ertur and Koch (2011) either, but, in general, firms maximize the present discounted value of future profits. However, since firms rent the services of the input factors labor and capital (in the form of intermediates), and there are no adjustment costs, no dynamic constraints exist, and the intertemporal maximization problem becomes a static one (or more precisely, a sequence of static problems (see e.g. Barro and Sala-i-Martin (2004, 32) and Acemoglu (2009, 435))).

[7] See Appendix B.1 for the derivation.

[8] See Appendix B.2 for the derivation.

availability of an increasing number of varieties. This effect, which can be interpreted as a form of technological progress has, for instance, been developed in the endogenous growth model by Romer (1990).[9]

The focus in the model described here is on technological interdependence (or more specifically, technology transfer) between regions. Hence, it is assumed that regions trade neither in goods nor in factors (Ertur and Koch, 2011, 220). Therefore, in general, the intermediate goods used and produced in region i as well as its final good are specific to this particular region. Nonetheless, due to technological interdependence, this is not the case for the process by which a specific intermediate good is produced. The respective idea for the production process might well have originated in a different region (Howitt, 2000, 831). The details of this idea will be provided in Section 3.

2.2 Intermediate Goods Sector

This section describes the production relations in the intermediate goods sector. It starts with the firms' optimization problem and illustrates the different assumptions underlying the generation of horizontal and vertical innovations. In general, horizontal innovations (or product innovations) increase the number of existing varieties, whereas vertical innovations (or process innovations) increase the productivity (quality) of an already existing variety.

2.2.1 Firms in the Intermediate Goods Sector

In the sector for intermediate goods, the production function for each monopolist in a given sector v is described by

$$x_i(v,t) = \frac{K_i(v,t)}{A_i(v,t)} \tag{5}$$

where $K_i(v,t)$ is the capital input in terms of the final good. From the functional form of the production function, it can be inferred that the production of varieties of higher quality becomes increasingly more capital intensive. This follows from the presence of the factor $A_i(v,t)$ in the denominator which rises with each new vintage of the good. In order to produce the intermediate good, the monopolist needs to rent capital at the price of $r_i(t) + \delta_i$ per unit, where $r_i(t)$ is the interest rate in region i and δ_i is the exogenously given region-specific depreciation rate. With this information, and, since $K_i(v,t)$ =

[9]The absence of an effect of an increasing number of varieties on productivity in the model presented here is demonstrated at the end of Appendix B.2.

$A_i(v,t)x_i(v,t)$ from Equation (5), it follows that the monopolist's profit function is

$$\pi_i(v,t) = p_i(v,t)x_i(v,t) - [r_i(t) + \delta_i] A_i(v,t)x_i(v,t). \tag{6}$$

Solving the inverse demand function in Equation (3) for $x_i(v,t)$ leads to the direct demand function for intermediates

$$x_i(v,t) = [\alpha A_i(v,t)]^{\frac{1}{1-\alpha}} l_i(t) p_i(v,t)^{-\frac{1}{1-\alpha}}. \tag{7}$$

Hence, the profit maximization problem for the monopolist is given by the constrained optimization problem of maximizing the profits in Equation (6) subject to the demand function in Equation (7). Substituting the expression for $x_i(v,t)$ into the profit function above, leads to the unconstrained profit maximization problem of the monopolist

$$\begin{aligned}\max_{p_i(v,t)} \pi_i(v,t) = {} & p_i(v,t) [\alpha A_i(v,t)]^{\frac{1}{1-\alpha}} l_i(t) p_i(v,t)^{-\frac{1}{1-\alpha}} \\ & - [r_i(t) + \delta_i] A_i(v,t) [\alpha A_i(v,t)]^{\frac{1}{1-\alpha}} l_i(t) p_i(v,t)^{-\frac{1}{1-\alpha}}.\end{aligned}$$

Setting the derivative $\frac{\partial \pi_i(v,t)}{\partial p_i(v,t)}$ equal to zero, results in the necessary condition

$$\begin{aligned}& -\frac{\alpha}{1-\alpha} [\alpha A_i(v,t)]^{\frac{1}{1-\alpha}} l_i(t) p_i(v,t)^{-\frac{\alpha}{1-\alpha}-1} \\ & + \frac{1}{1-\alpha} [r_i(t) + \delta_i] \alpha^{\frac{1}{1-\alpha}} A_i(v,t)^{\frac{2-\alpha}{1-\alpha}} l_i(t) p_i(v,t)^{-\frac{1}{1-\alpha}-1} = 0\end{aligned}$$

and solving for the profit-maximizing price yields

$$p_i(v,t) = [r_i(t) + \delta_i] \frac{A_i(v,t)}{\alpha}. \tag{8}$$

Substituting this price into Equation (7) leads to

$$x_i(v,t) = \alpha^{\frac{2}{1-\alpha}} l_i(t) [r_i(t) + \delta_i]^{-\frac{1}{1-\alpha}}.$$

This result shows that the production of the intermediate good is independent of v (i.e. independent of the specific variety produced), and hence it holds that

$$x_i(v,t) = x_i(t), \tag{9}$$

implying that the equilibrium in the intermediate goods sector is symmetric so that independent of the specific variety v all monopolists produce the identical amount $x_i(t)$

of their respective variety.[10]

Noting that in equilibrium $x_i(t) = \hat{k}_i(t)l_i(t)$ holds,[11] it follows that the equilibrium interest rate is given by

$$r_i(t) = \alpha^2 \hat{k}_i(t)^{\alpha-1} - \delta_i. \tag{10}$$

Finally, using the profit-maximizing price in Equation (8), substituting the equilibrium interest rate and the expression for the quantity, $x_i(t) = \hat{k}_i(t)l_i(t)$, in the symmetric equilibrium into the profit function in Equation (6), implies that the monopolist's profits are given by

$$\pi_i(v,t) = A_i(v,t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t), \tag{11}$$

where the function $\tilde{\pi}_i\big(\hat{k}_i(t)\big)$ is defined as $\tilde{\pi}_i\big(\hat{k}_i(t)\big) \equiv \alpha(1-\alpha)\hat{k}_i(t)^{\alpha}$.

2.2.2 Horizontal Innovations in the Intermediate Goods Sector

The relevant assumption concerning horizontal innovations is that new varieties are created by imitation. Moreover, no resources are spent on this activity so that imitation is not a deliberate effort by individuals. As Aghion and Howitt (1998, 107) laconically put it: "imitation just happens". Hence, individuals in the economy can be sure that new varieties will enter the economy, but the specific point in time when a new intermediate good will be available for production of final output or when a new sector opens up in which to reap monopoly profits remains uncertain. Therefore, the occurrence of innovations is governed by a random process, and the specific random process assumed is a Poisson process. In more formal terms, each agent in region i imitates with equal likelihood, and her Poisson arrival rate[12] of imitation is given by $\xi > 0$, which is identical across regions. This implies that the aggregate flow of new intermediate goods is given by

$$\dot{Q}_i(t) = \xi L_i(t). \tag{12}$$

[10] Naturally, this also results, if one sets up the profit maximization problem with quantity as the decision variable (see, for example, Varian, 1992, 234) so that

$$\max_{x_i(v,t)} = \alpha A_i(v,t)x_i(v,t)^{\alpha-1}l_i(t)^{1-\alpha}x_i(v,t) - [r_i(t)+\delta_i]\,A_i(v,t)x_i(v,t).$$

Taking the derivative with respect to quantity, it follows that the marginal revenue and marginal cost function are proportional to $A_i(v,t)$, and, since this is the only difference between the firms producing an intermediate product, the symmetric equilibrium in Equation (9) follows (Howitt, 2000, 832).

[11] This result is derived in Appendix B.2 as an intermediate result in the derivation of the production function in intensive form.

[12] See Appendix A for a primer on Poisson processes.

As Appendix B.3 demonstrates, the number of workers per variety $l_i(t)$ converges to

$$l_i = \frac{n_i}{\xi}, \tag{13}$$

which is independent of time t and thus constant.

2.3 Research and Development – Vertical Innovations

Apart from increases in the number of intermediate goods (horizontal innovations), a key characteristic of the model are quality, i.e. productivity, improvements of already existing intermediate products (vertical innovations). On a general level, quality improvements for a given variety result from investment in R&D in that particular sector.[13] Here, the final good is the relevant input factor. It is assumed that the inventor of a higher-quality variety in sector v at the same time also is the producer of this intermediate good.[14]

The mere fact of engaging in research activities naturally is no guarantee for success. As is standard in this type of models (see, for example, Aghion and Howitt, 1998, 54-55), the underlying random process for the occurrence of vertical innovations is also assumed to be a Poisson process. However, in this case, the Poisson arrival rate in any sector $v \in [0, Q_i(t)]$ is slightly more complicated as it is not given by a single parameter, but instead by the function

$$\phi_i(t) = \lambda_i \kappa_i(t)^{\phi}. \tag{14}$$

The variable $\kappa_i(t)$ denotes the sector-specific expenditures on vertical R&D adjusted for productivity, and the parameter ϕ, for which $0 \leq \phi \leq 1$ holds, gauges the strength of a given amount of R&D expenditures on λ_i (Ertur and Koch, 2011, 222). To be more precise with respect to R&D expenditures, these are given by $\kappa_i(t) = \frac{S_{A,i}(t)}{Q_i(t)A_i(t)^{max}}$, where $S_{A,i}(t)$ is the total input into R&D in region i, so that $\frac{S_{A,i}(t)}{Q_i(t)}$ reflects the total amount invested in a given sector aggregated over all firms. $A_i(t)^{max}$ is the maximal value of $A_i(v,t)$ (or the leading-edge productivity parameter), and it is defined by

$$A_i(t)^{max} \equiv \max \left\{ A_i(v,t); v \in [0, Q_i(t)] \right\}. \tag{15}$$

[13] The specific setup in the intermediate sector with imitation leading to new varieties and innovation to a higher quality of existing varieties was introduced by Young (1998), who formalized ideas expressed verbally in earlier work by Gilfillan (1935a,b). In this approach, a scale effect (i.e. a positive effect of population on the per capita growth rate), which was criticized by Jones (1995a,b) is not present. See also, Aghion and Howitt (1998, 106-110).

[14] This assumption is made for convenience. As Barro and Sala-i-Martin (2004, 290) state, results would be the same, if one alternatively assumed that inventors charged producers of intermediate goods a license fee for the use of the blueprint or process innovation.

An important assumption is made concerning this parameter. Potential innovators all have immediate access to this technological knowledge and thus "all draw on the same pool" (Aghion and Howitt, 1998, 87-88).

Adjustment of the sector-specific resource investment by the leading-edge technology parameter captures the assumption of ever increasing complexity in the research process (Ertur and Koch, 2011, 222). With technology ever increasing, more and more resources need to be spent to prevent the rate of innovation from slowing down. In other words, "as technology advances, the resource cost of further advances increases proportionally" (Aghion and Howitt, 1998, 410). Note that, since the prospective payoffs from an innovation are identical across sectors, productivity-adjusted R&D investment, $\kappa_i(t)$, is also identical for each sector in region i.

Potential innovators face the questions of whether to conduct research at all, and if so how much to invest in R&D. Concerning these decisions, the value of an innovation to a successful innovator in a given sector is a critical variable. This value is given by

$$V_i(t) = \int_t^{\infty} e^{-\int_t^{\tau}(r_i(s)+\phi_i(s))ds} \pi_i(\tau)\, d\tau. \tag{16}$$

At some point in the future, a higher-quality variety will be invented in this sector, and the incumbent will be replaced by the successful innovator and lose his profits.[15] The equation above takes this into account and adjusts for it by including the Poisson arrival rate of new innovations in the discount factor.[16] Adjusted for productivity, the value of an innovation is defined as $v_i(t) \equiv \frac{V_i(t)}{A_i(t)^{max}}$ (Ertur and Koch, 2011, 222) so that

$$v_i(t) = \int_t^{\infty} e^{-\int_t^{\tau}(r_i(s)+\phi_i(s))\, ds} \frac{1}{A_i(\tau)^{max}} \pi_i(\tau)\, d\tau.$$

Substituting for $\pi_i(\tau)$ from Equation (11) and noting that by assumption the productivity level of a firm that innovates at time t is at the leading edge, implies that $A_i(v,t) =$

[15]Innovations will result from new entrants into the sector due to the Arrow replacement effect (Arrow, 1962). This effect states that incumbents who innovate would only replace part of their existing profits. On the other hand, researchers entering the sector have access to the leading-edge technology parameter, and, if they are successful, can reap the complete monopoly profits. Hence, these researchers have higher incentives to innovate than incumbents.

[16]A formal derivation of this value is provided in Appendix B.4.

$A_i(t)^{max}$ in this case. Therefore[17]

$$v_i(t) = \int_t^{\infty} e^{-\int_t^{\tau}(r_i(s)+\phi_i(s))\,ds} \frac{A_i(\tau)^{max}}{A_i(\tau)^{max}} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\, d\tau$$
$$= \int_t^{\infty} e^{-\int_t^{\tau}(r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\, d\tau.$$

Taking the derivative of this equation with respect to time, leads to the following research-arbitrage equation[18]

$$\frac{\dot{v}_i(t)}{v_i(t)} = r_i(t) + \phi_i(t) - \frac{l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)}{v_i(t)}. \tag{17}$$

Written in this form, the function in Equation (16) is also known as the (stationary) Hamilton-Jacobi-Bellman Equation (see, for example, Acemoglu, 2009, 245 and 462-463). Expressed equivalently as

$$r_i(t)v_i(t) = l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big) + \dot{v}_i(t) - \phi_i(t)v_i(t),$$

it shows that the required return on an innovation, $r_i(t)v_i(t)$, for a firm that engages in R&D, needs to equal its flow profits, $l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)$, plus any capital gains, $\dot{v}_i(t)$, adjusted for the fact that with positive probability $\phi_i(t)$ a new innovation occurs at some point in time, and the monopolist's product thus becomes obsolete from this point onwards.

An individual considering conducting R&D with the aim of improving a particular variety v has expected profits $\pi^e_{A,i}$. In particular,

$$\pi^e_{A,i} = \lambda_i\kappa_i(t)^{\phi} \frac{S_{A,i}(v,t)}{S_{A,i}(t)/Q_i(t)} \cdot V_i(t) + (1-\lambda_i\kappa_i(t)^{\phi}) \frac{S_{A,i}(v,t)}{S_{A,i}(t)/Q_i(t)} \cdot 0 - S_{A,i}(v,t). \tag{18}$$

Here, $\lambda_i\kappa_i(t)^{\phi}$ is the probability of being successful in research, and $1-\lambda_i\kappa_i(t)^{\phi}$ is the complementary probability of failure in research. $S_{A,i}(v,t)$ denotes how many resources the firm invests in R&D, and the division by $S_{A,i}(t)/Q_i(t)$ captures negative externalities in the research process. More precisely, overlap and duplication of research efforts are underlying this assumption (Ertur and Koch, 2011, 222). Hence, there is no linear increase in profits with resources invested in R&D. Note that the R&D technology requires only output as an input.[19] In other words, only laboratory equipment is required to engage in research activities, but no workers or scientists need to be employed. Therefore, this

[17]In the article by Ertur and Koch the dependence of the number of workers per variety l_i on τ is missing.

[18]This derivation involves applying Leibniz's Formula, and the detailed steps are provided in Appendix B.5.

[19]The price of these resources is normalized to 1 as they are measured in units of the output good, which is the numéraire in this model.

model is a variant of a "lab-equipment" model (see, for instance, Acemoglu, 2009, 433).

Incumbent firms in the R&D sector then face the following profit-maximization problem (this follows from simplifying Equation (18) and dropping the superscript for expectations to enhance readability)

$$\max_{S_{A,i}(v,t)} \pi_{A,i}(v) = \lambda_i \kappa_i(t)^{\phi} \frac{S_{A,i}(v,t)}{S_{A,i}(t)/Q_i(t)} V_i(t) - S_{A,i}(v,t).$$

The necessary condition therefore reads

$$\frac{\partial \pi_{A,i}(v)}{\partial S_{A,i}(v,t)} = 0 \qquad \Longleftrightarrow \qquad \frac{\lambda_i \kappa_i(t)^{\phi}}{S_{A,i}(t)/Q_i(t)} V_i(t) = 1.$$

By substituting $V_i(t) = v_i(t) A_i(t)^{max}$ and using the definition of $\kappa_i(t)$ in this condition, it follows that the value of an innovation is given by

$$v_i(t) = \frac{1}{\lambda_i} \kappa_i(t)^{1-\phi}.$$

Solving for $\kappa_i(t)$ and log-differentiating the resulting expression yields $\frac{\dot{\kappa}_i(t)}{\kappa_i(t)} = \frac{1}{1-\phi} \frac{\dot{v}_i(t)}{v_i(t)}$. Substituting thereafter from the research-arbitrage equation in (17) and then inserting the expression for the Poisson arrival rate from (14) leads to the following differential equation

$$\frac{\dot{\kappa}_i(t)}{\kappa_i(t)} = \frac{1}{1-\phi} \left[r_i(t) + \lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi-1} l_i(t) \tilde{\pi}_i\big(\hat{k}_i(t)\big) \right]. \tag{19}$$

This equation describes how the resources invested in R&D (measured in terms of the final good) evolve over time.

In the derivation of this expression, the leading-edge productivity parameter $A_i(t)^{max}$ has been used. As innovations result in knowledge spillovers, this parameter is not constant. In particular, its growth rate and thereby the growth rate of technological progress is equal to

$$g_i(t) \equiv \frac{\dot{A}_i(t)^{max}}{A_i(t)^{max}} = \frac{\sigma}{Q_i(t)} Q_i(t) \lambda_i \kappa_i(t)^{\phi} = \sigma \lambda_i \kappa_i(t)^{\phi}. \tag{20}$$

Basically, therefore, $A_i(t)^{max}$ grows with the aggregate rate of innovations (i.e. the Poisson arrival rate from Equation (14) times the number of differentiated varieties $Q_i(t)$) multiplied by a factor of proportionality $\sigma/Q_i(t) > 0$. This factor captures by how much public knowledge increases as a result of an additional innovation or, expressed differently, it measures "the marginal impact of each innovation on the stock of public knowledge" (Aghion and Howitt, 1998, 411). However, this impact is diminishing in $Q_i(t)$. Over time, horizontal innovations lead to an increase in the number of intermediates, and the division of the factor of proportionality by this number ensures that innovations of a given size for a particular product, will have a diminishing impact (Ertur and Koch, 2011, 223).

Having determined the growth rate for the leading-edge productivity parameter, it is helpful for subsequent derivations to look at the corresponding growth rate for the average productivity parameter, $A_i(t)$. In general, a successful innovation for intermediate good v changes productivity for this good from $A_i(v,t)$ to $A_i(t)^{max}$.[20] Across innovating sectors, the average increase from a successful innovation is given by $A_i(t)^{max} - A_i(t)$. Taking into account that innovations are generated with rate $\lambda_i \kappa_i(t)^{\phi}$ uniformly across all sectors, and that average productivity remains unaffected by horizontal innovations, it follows that the change in average productivity can be expressed as

$$\dot{A}_i(t) = \lambda_i \kappa_i(t)^{\phi} \big(A_i(t)^{max} - A_i(t) \big).$$

Appendix B.7 demonstrates that the ratio of the leading-edge productivity parameter to the average productivity parameter converges to the constant $1+\sigma$ so that $A_i(t)^{max} = (1+\sigma)A_i(t)\,\forall\, t$, implying that the growth rates of both variables will be identical.

2.4 Physical Capital Accumulation and Steady State

As in a standard neoclassical Solow model, the accumulation of physical capital is governed by the general equation

$$\dot{\hat{k}}_i(t) = s_{K,i}\hat{k}_i(t)^{\alpha} - \big(n_i + g_i(t) + \delta_i\big)\hat{k}_i(t). \tag{21}$$

Here, $s_{K,i}$ denotes the investment rate for physical capital in region i and δ_i signifies the depreciation rate for physical capital, which is region-specific. The evolution of the economy can then be described by the following system of differential equations:

$$\begin{aligned}
\dot{\hat{k}}_i(t) &= s_{K,i}\hat{k}_i(t)^{\alpha} - \big(n_i + g_i(t) + \delta_i\big)\hat{k}_i(t) \\
\dot{\kappa}_i(t) &= \frac{\kappa_i(t)}{1-\phi}\Big[r_i(t) + \lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi-1} l_i(t) \tilde{\pi}_i\big(\hat{k}_i(t)\big) \Big]
\end{aligned}$$

where the first equation follows from Equation (21) by inserting for the growth rate from Equation (20), and the second equation above is just Equation (19) multiplied by $\kappa_i(t)$. In steady state, capital in efficiency units and productivity-adjusted R&D investment are

[20] One might wonder about the distribution of productivities across sectors in this model. Appendix B.6 demonstrates that the relative productivities $a_i(v,t) = A_i(v,t)/A_i(t)^{max}$ converge to an invariant distribution, meaning that even though $A_i(t)^{max}$ increases over time and sectors change position in the distribution, its shape remains constant in the long run (Aghion and Howitt, 1992, 88).

constant so that $\dot{\hat{k}}_i(t) = \dot{\kappa}_i(t) = 0$. Imposing this condition and denoting steady-state values with an asterisk, implies that the steady-state rate of technological progress in region i is given by $g_i^* = \sigma\lambda_i \left(\kappa_i^*\right)^\phi$ and that the steady-state value for $\hat{k}_i^*$ is defined by the $\dot{\hat{k}}_i(t) = 0$-isocline as[21]

$$\hat{k}_i^* = \left(\frac{s_{K,i}}{n_i + \sigma\lambda_i \left(\kappa_i^*\right)^\phi + \delta_i}\right)^{\frac{1}{1-\alpha}}. \tag{22}$$

This isocline is depicted as the downward-sloping curve (I) in $(\kappa_i(t) - \hat{k}_i(t))$-space in the upper right hand in Figure 3. From setting $\dot{\kappa}_i(t) = 0$, it follows that the $\dot{\hat{k}}_i(t) = 0$-isocline is given by

$$1 = \lambda_i \left(\kappa_i^*\right)^{\phi-1} \frac{\tilde{\pi}_i(\hat{k}_i^*) l_i}{r_i^* + \lambda_i \left(\kappa_i^*\right)^\phi}.$$

This relation is the upward-sloping schedule labeled (II) in Figure 3. Curve (I) is downward sloping as in steady state an increase in R&D investment leads to an increase in the growth rate g_i^*. From Equation (22), it then follows that for equilibrium to be maintained the capital-output ratio, $\frac{\hat{k}^*}{\hat{y}^*} = (\hat{k}^*)^{1-\alpha}$, needs to fall. On the other hand, Curve (II) is upward sloping, since when $\hat{k}_i^*$ increases, the interest rate in steady state falls (compare Equation (10)) and profits increase (see Equation (11)). Hence, in equilibrium R&D expenditures need to rise.

Turning now to the remaining parts of Figure 3, the lower right one shows the Solow diagram as, for example, in Barro and Sala-i-Martin (2004, 56). The main difference to the standard version is that here the effective depreciation rate, $n_i + g_i(t) + \delta_i$, through its dependence on the rate of technological progress, $g_i(t)$, is endogenously determined by investment in R&D and thus moves up until the steady state is reached (Ertur and Koch, 2011, 224). This determination of $g_i(t)$ through $\kappa_i(t)$ is depicted in the upper left part of the figure, whereas the positive dependence of the effective depreciation rate on technological progress is depicted in the lower left part of the figure. In steady state, with $g_i(t) = g_i^*$, the effective depreciation rate is constant, which allows for determining the level of physical capital per effective worker and the level of R&D investment via the dotted lines.

[21] There seems to be a typo in the corresponding Equation (22) in Ertur and Koch (2011), where the left-hand side should read $(\hat{k}_i^*)^{1-\alpha}$ instead of $(\hat{k}_i^*)^\alpha$.

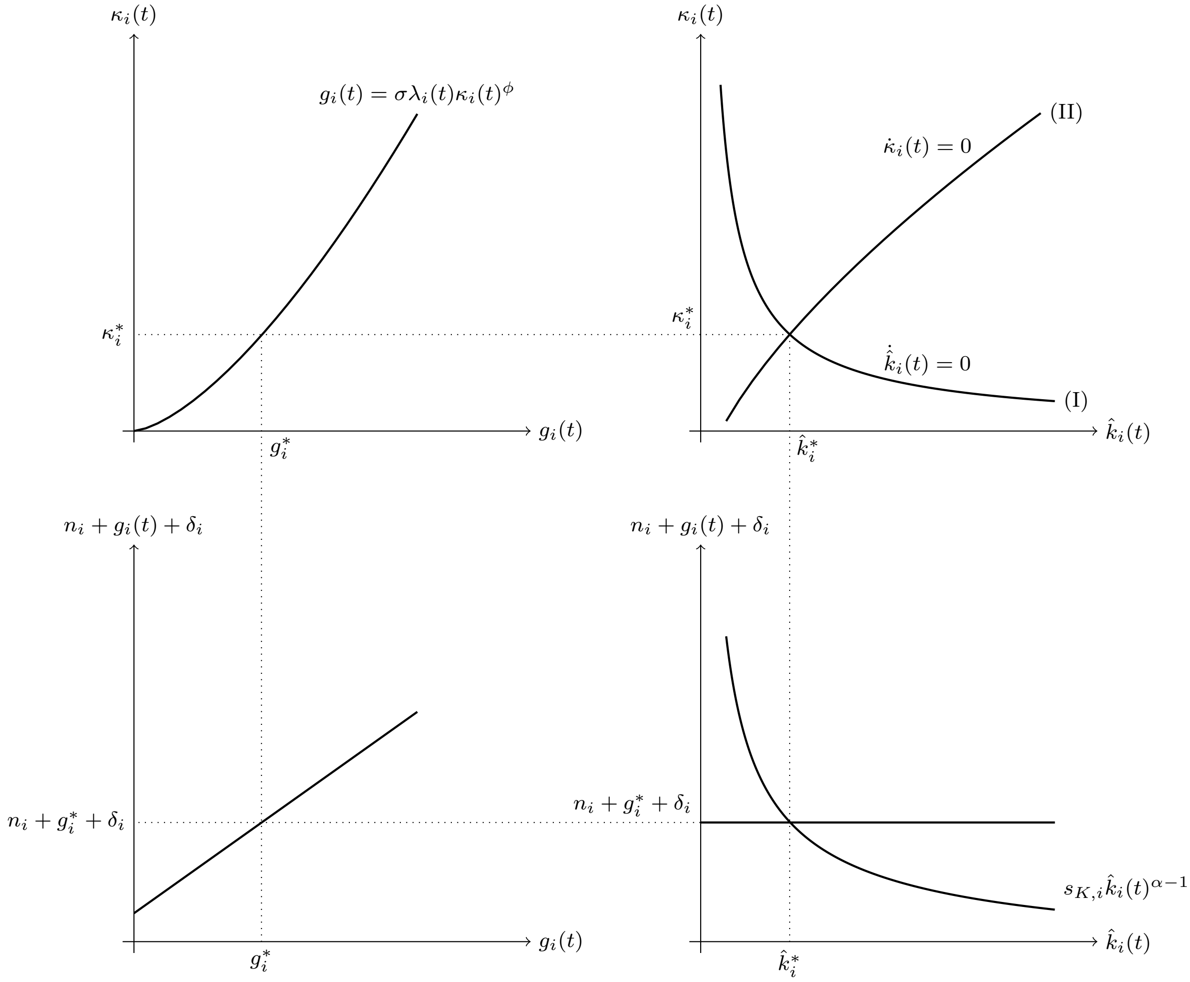


Figure 3: Illustration of the Steady State (Adapted from Ertur and Koch (2011)).

3 Multi-Region Schumpeterian Growth Model with Technological Interdependence

This section introduces the analytical setup in which diffusion of knowledge depends on a region's gap to its own technological frontier. In addition, the steady-state equation on which the estimation will be based, is derived.

3.1 Research Productivity, Knowledge Spillovers, and Technology Gap

Turning now to the case of multiple regions, the assumption that all regions develop independently from each other is abandoned. Interdependence enters the model via the assumption that the productivity in the research sector, λ_i, in region i depends on its own level of technology relative to the level of other regions as well as on the way the connection between regions is modeled. In formal terms, the region-specific research productivity is given by

$$\lambda_i = \lambda \prod_{j=1}^{N} \left(\frac{A_j(t)}{A_i(t)} \right)^{\gamma_i v_{ij}} . \tag{23}$$

Note that the technology frontier is specific to each region due to the presence of the parameters v_{ij}. Concerning these, it is assumed that they are non-negative, finite and non-stochastic. Moreover, $\sum_{j=1}^{N} v_{ij} = 1$ is assumed. In general, not all regions necessarily are equally able to increase their research productivity due to a given increase in knowledge in the regions it is connected to. In this regard, the absorptive capacity of a region plays an important role.[22] This notion is picked up by Ertur and Koch (2011) in the parameter γ_i, as it is assumed that the absorption capacity depends on the human capital stock, H_i, in region i in the following way: $\gamma_i = \gamma H_i$, where $\gamma < 1$ is a measure of the amount of knowledge spilling over from other regions. At this point, the derivation in Appendix B.7, which demonstrates that the growth rates of the leading-edge productivity parameter and the average productivity parameter are identical, becomes helpful. Substituting the expression for λ_i into Equation (20), and using that $g_i(t) \equiv \frac{\dot{A}_i(t)^{max}}{A_i(t)^{max}} = \frac{\dot{A}_i(t)}{A_i(t)}$ leads to

$$g_i(t) \equiv \sigma \lambda \kappa_i(t)^{\phi} \prod_{j=1}^{N} \left(\frac{A_j(t)}{A_i(t)} \right)^{\gamma_i v_{ij}} .$$

The last term in this equation represents the distance to the technological frontier for region i. This implies that the further away a region is from its own technology frontier, i.e. the larger is the average technological level in the regions it is connected to or the lower is its own level of technology, the higher is its productivity in the research sector. The intuition is that there exists a large pool of knowledge in the region's environment into which it has not yet tapped into. Spillovers from other regions or equivalently spatial externalities are comparatively large in this case.[23] Conversely, a region close to its technological frontier cannot benefit from spillovers or technology diffusion from connected regions in the same extent as the pool of knowledge has been largely tapped out and copying "foreign" technology becomes more difficult (Ertur and Koch, 2011, 226).[24]

Since in steady state $\hat{k}_i$ and κ_i grow at constant rates in each region, it follows that a

[22]This corresponds to ideas developed in Nelson and Phelps (1966), although the specific word "absorptive capacity" is not mentioned by them.

[23]As Ertur and Koch (2011, 217-218) point out, this is the concept of the "advantage of backwardness" by Gerschenkron (1962).

[24]These effects are similar to the effects of "standing on the shoulders of giants" (compare Caballero and Jaffe, 1993) and "fishing out" (see Jones, 1995a, 765) mentioned in the literature on endogenous growth models with respect to the research productivity in a single country.

region's distance to its own technological frontier remains constant. However, for steady state to occur this requires that all regions grow at identical rates or, expressed differently, converge to parallel growth paths in the long run. This steady state growth rate for regions $i = 1, \dots, N$ is given by

$$g^w \equiv \sigma\lambda\kappa_i^\phi \prod_{j=1}^{N} \left(\frac{A_j}{A_i}\right)^{\gamma_i v_{ij}} \tag{24}$$

where time dependence t as well as the asterisks indicating steady-state values have been dropped to enhance readability. Regions converge to the same growth rate in the long run due to the inverse relation between how many resources are invested in the research sector and this sector's productivity in steady state. Investing a comparatively large amount of resources in the research sector so that κ_i is relatively high, implies that the level of technology will in turn also be relatively high. From Equation (23) it then follows that the ratio of the average level of technology to the own level of technology will be comparatively low, i.e. a region is close to its own technology frontier, which implies that research productivity λ_i in turn will be relatively low, too. A region with comparatively low R&D expenditures has a relatively high research productivity due to the large distance to its own technology frontier and as Ertur and Koch (2011, 226) note, due to technology diffusion and its impact on research productivity, convergence to the steady state growth rate occurs.

In order to test the model empirically in the following section, Equation (24) will be rewritten. As an intermediate step note that the productivity-adjusted sector-specific expenditures into R&D are given by $\kappa_i(t) = \frac{S_{A,i}(t)}{Q_i(t)A_i(t)^{max}}$. Multiplying and dividing this expression by $\frac{Y_i}{L_i}$ and using $A_i(t)^{max} = (1+\sigma)A_i(t)$ from Appendix B.7 leads to $\kappa_i = \frac{S_{A,i}}{Y_i}\frac{Y_i}{L_i}\frac{L_i}{Q_i}\frac{1}{(1+\sigma)A_i}$. With the result from Equation (13), this can be equivalently expressed as

$$\kappa_i = s_{A,i} y_i \frac{n_i}{\xi} \frac{1}{(1+\sigma)A_i}. \tag{25}$$

Here, the definition $s_{A,i} \equiv \frac{S_{A,i}}{Y_i}$ for the investment rate in the research sector has been applied. The global technology growth rate can then be shown to be given by the expression[25]

$$g^w = \frac{\sigma\lambda}{[(1+\sigma)\xi]^\phi} s_{A,i}^\phi y_i^\phi n_i^\phi A_i^{-\phi-1} \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}}. \tag{26}$$

[25]The derivation might not be immediately obvious and is therefore given in Appendix B.8.

Applying the natural logarithm to this equation and then solving for $\ln A_i$ yields

$$\ln A_i = \frac{1}{1+\phi} \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} + \frac{\phi}{1+\phi} (\ln s_{A,i} + \ln n_i + \ln y_i) + \frac{\gamma H_i}{1+\phi} \sum_{j \neq i}^{N} v_{ij} \ln A_j.$$

Stacking the equations for regions $i = 1, \ldots, N$, the level of technology is given by

$$\underbrace{\begin{pmatrix} \ln A_{1t} \\ \vdots \\ \ln A_{Nt} \end{pmatrix}}_{\substack{=\boldsymbol{A} \\ (N\times 1)}} = \frac{1}{1+\phi} \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \boldsymbol{\iota} + \frac{\phi}{1+\phi} \underbrace{\begin{pmatrix} \ln s_{A,1} + \ln n_1 + \ln y_1 \\ \vdots \\ \ln s_{A,N} + \ln n_N + \ln y_N \end{pmatrix}}_{\substack{=\boldsymbol{s_A}+\boldsymbol{n}+\boldsymbol{y} \\ (N\times 1)}}$$

$$+ \frac{\gamma}{1+\phi} \underbrace{\begin{pmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & H_N \end{pmatrix}}_{\substack{=\boldsymbol{H}=diag(H_i) \\ (N\times N)}} \underbrace{\begin{pmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NN} \end{pmatrix}}_{\substack{=\boldsymbol{V} \\ (N\times N)}} \underbrace{\begin{pmatrix} \ln A_1 \\ \vdots \\ \ln A_N \end{pmatrix}}_{\substack{=\boldsymbol{A} \\ (N\times 1)}}$$

Defining $\boldsymbol{W} \equiv \boldsymbol{H} \cdot \boldsymbol{V}$ with entries $v_{ii} = 0$ if $i = j$ in $\boldsymbol{V}$, the equivalent matrix expression for the level of technology is

$$\boldsymbol{A} = \frac{1}{1+\phi} \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \boldsymbol{\iota} + \frac{\phi}{1+\phi} (\boldsymbol{s_A} + \boldsymbol{n} + \boldsymbol{y}) + \frac{\gamma}{1+\phi} \boldsymbol{W}\boldsymbol{A}. \tag{27}$$

Given that the matrix $\left(\boldsymbol{I} - \frac{\gamma}{1+\phi}\boldsymbol{W}\right)$ is non singular and thus has an inverse,[26] Equation (27) can be solved for $\boldsymbol{A}$ to yield a matrix equation for the level of technology

$$\begin{aligned} \boldsymbol{A} = &\frac{1}{1+\phi} \left(\boldsymbol{I} - \frac{\gamma}{1+\phi}\boldsymbol{W}\right)^{-1} \left(\ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \boldsymbol{\iota}\right) \\ &+ \frac{\phi}{1+\phi} \left(\boldsymbol{I} - \frac{\gamma}{1+\phi}\boldsymbol{W}\right)^{-1} (\boldsymbol{s_A} + \boldsymbol{n} + \boldsymbol{y}). \end{aligned} \tag{28}$$

3.2 Income per Worker in Steady State

In this section an expression that determines the income per worker in steady state will be derived. From Equation (4) it follows that the production function per worker in steady state for region i is given by $y_i^* = A(\hat{k}_i^*)^{\alpha}$. Substituting for the steady-state level

[26] An application of Gerschgorin's Theorem (see Gerschgorin, 1931) ensures that. See Appendix B.9 for a similar case.

of capital in efficiency units leads to

$$y_i^* = A \left(\frac{s_{K,i}}{n_i + \sigma\lambda_i \left(\kappa_i^*\right)^{\phi}} + \delta_i \right)^{\frac{1}{1-\alpha}}.$$

After taking the natural logarithm and stacking the expressions for regions $i = 1, \ldots, N$, the steady-state incomes in per worker terms can be expressed in the following matrix equation: $\boldsymbol{y} = \boldsymbol{A} + \frac{\alpha}{1-\alpha}\boldsymbol{s_K}$ in which the matrix $\boldsymbol{s_K}$ is an $N \times 1$ matrix with the terms $\frac{s_{K,i}}{n_i+g^w+\delta_i}$ for the respective regions. Inserting the result for $\boldsymbol{A}$ from Equation (28) into the expression above, yields

$$\boldsymbol{y} = \left(\ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \right) \boldsymbol{\iota} + \phi(\boldsymbol{s_A} + \boldsymbol{n}) + \frac{\alpha(1+\phi)}{1-\alpha}\boldsymbol{s_K} - \frac{\alpha\gamma}{1-\alpha}\boldsymbol{W}\boldsymbol{s_K} + \gamma\boldsymbol{W}\boldsymbol{y}. \quad (29)$$

Writing this equation for an individual region i clarifies the determinants of the level of per worker income in steady state

$$\begin{aligned} \ln y_i = \; & \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} + \phi(\ln s_{A,i} + \ln n_i) + \frac{\alpha(1+\phi)}{1-\alpha} \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} \\ & - \frac{\alpha\gamma H_i}{1-\alpha} \sum_{j \neq i}^{N} v_{ij} \ln \frac{s_{K,j}}{n_j + g^w + \delta_j} + \gamma H_i \sum_{j \neq i}^{N} v_{ij} \ln y_j. \end{aligned} \quad (30)$$

It is important to note here that a change in the independent variables in region i affects the steady-state levels in the regions to which it is connected, and the steady-state levels in neighboring regions in turn have an influence on the respective level in region i. Therefore, studying the effect of, for example, a change in the investment rate in physical capital requires an analysis of the complete interdependent system in Equation (29). In general, the impact of a change in one of the independent variables can be divided into two parts. The first one represents the impact on the income per worker in steady state in region i due to a change in the independent variable in this region, and the second one details the effect of an identical change in the same variable in all regions $j = 1, \ldots, N$ with $j \neq i$ that region i is connected to. For example, the $N \times N$ matrix of income per worker elasticities with respect to the R&D investment rate $\boldsymbol{s_A}$, is given by

$$\boldsymbol{\eta}^{\boldsymbol{s_A}} \equiv \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{s_A}} = \phi(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1} = \phi\boldsymbol{I} + \phi\sum_{r=1}^{\infty}\gamma^r\boldsymbol{W}^r. \quad (31)$$

This result is obtained by solving Equation (29) for $\boldsymbol{y}$ and then differentiating the result with respect to $\boldsymbol{s_A}$.[27] Concerning the last equality, it follows as the inverse $(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}$ is given by the Neumann series $\sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r$ (see, for instance, Meyer (2000, 126 and 618) for this result) so that

$$(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1} = \boldsymbol{I} + \gamma \boldsymbol{W} + \gamma^2 \boldsymbol{W}^2 + \cdots + \gamma^r \boldsymbol{W}^r + \cdots = \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r. \tag{32}$$

This series is also called the spatial multiplier.[28] With respect to the elasticity in Equation (31), it highlights that changes in R&D investment in a given region i will have an impact on income per worker in all other locations. Hence, the total effect can be decomposed into the two impacts described above.

The first effect is given by[29]

$$\eta_i^{s_{A,i}} = \phi + \phi \sum_{r=1}^{\infty} \gamma_i^r v_{ii}^{(r)} > 0 \tag{33}$$

where $v_{ii}^{(r)}$ denotes the element i in row i and column i of the matrix $\boldsymbol{V}$ taken to the power of r. The second effect, the impact on region i of a change in R&D expenditures in the regions it is connected to, is

$$\eta_i^{s_{A,j}} = \phi \sum_{r=1}^{\infty} \gamma_i^r v_{ij}^{(r)} > 0. \tag{34}$$

In a similar manner, the aggregate effect of changes in the physical capital investment rate can be derived to yield

$$\boldsymbol{\eta}^{\boldsymbol{s_K}} \equiv \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{s_K}} = \frac{\alpha}{1-\alpha}\boldsymbol{I} + \frac{\alpha\phi}{1-\alpha}(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1} = \frac{\alpha(1+\phi)}{1-\alpha}\boldsymbol{I} + \frac{\alpha\phi}{1-\alpha}\sum_{r=1}^{\infty} \gamma^r \boldsymbol{W}^r \tag{35}$$

[27] Naturally, this derivation is only valid given that the inverse $(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}$ exists. Appendix B.9 provides the conditions under which this inverse exists.

[28] See e.g. Ertur and Koch (2011, 232), Elhorst (2010, 21-22), or LeSage and Pace (2014) on this expression.

[29] Note that even though the entries v_{ii} in the matrix $\boldsymbol{V}$ might be zero, this is not necessarily the case for entries in the corresponding matrix raised to a higher order as the following counterexample shows:

$$\boldsymbol{V}\boldsymbol{V} = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} = \begin{pmatrix} 4/9 & 1/9 & 3/9 & 1/9 \\ 1/6 & 1/3 & 1/6 & 1/3 \\ 1/3 & 1/9 & 4/9 & 1/9 \\ 1/6 & 1/3 & 1/6 & 1/3 \end{pmatrix} = \boldsymbol{V}^2$$

Therefore, the matrix $\boldsymbol{V}$ is not idempotent. Economically, this effect can be understood as knowledge spilling over from region i to region j from where a spillover originates back to region i. In other words, feedback effects exist in this model.

which is positive, as knowledge diffuses across regions. For the employment growth rate, the corresponding elasticity is given by

$$\begin{aligned}\boldsymbol{\eta}^n \equiv \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{n}} = & -\frac{\alpha}{1-\alpha}\mathbf{diag}\left(\frac{\boldsymbol{n}}{\boldsymbol{n}+\boldsymbol{g}+\boldsymbol{\delta}}\right) + \frac{\alpha\phi}{1-\alpha}\mathbf{diag}\left(\frac{\boldsymbol{g}+\boldsymbol{\delta}}{\boldsymbol{n}+\boldsymbol{g}+\boldsymbol{\delta}}\right) \\ & + \frac{\alpha\phi}{1-\alpha}\sum_{r=1}^{\infty}\gamma^r \boldsymbol{W}^r \mathbf{diag}\left(\frac{\boldsymbol{g}+\boldsymbol{\delta}}{\boldsymbol{n}+\boldsymbol{g}+\boldsymbol{\delta}}\right).\end{aligned} \tag{36}$$

This elasticity captures that on the one hand, per worker income is positively influenced by increases in the employment growth rate, as this leads to a larger number of horizontally differentiated products on which R&D can be conducted, and it captures that on the other hand, a negative impact exists, which results from the dilution of physical capital (Ertur and Koch, 2011, 250).[30]

4 Empirical Specification and Estimation Method

This section describes the empirical specification of the model and details the econometric estimation method. In particular, the derivation of the log-likelihood function and its concentrated version will be discussed in detail.

4.1 Empirical Specification

From the expression for the steady-state level of income per worker in Equation (30), the following empirical counterpart in reduced form can be derived[31]

$$\begin{aligned}\ln y_i = & \beta_0 + \beta_1 \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \beta_2 \ln s_{A,i} + \beta_3 \ln n_i \\ & + \theta H_i \sum_{j\neq i}^{N} v_{ij} \ln \frac{s_{K,j}}{n_j + g^w + \delta_j} + \gamma H_i \sum_{j\neq i}^{N} v_{ij} \ln y_j + \varepsilon_i.\end{aligned} \tag{37}$$

In this equation, the parameters are given by the following expressions $\beta_0 \equiv \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^\phi} > 0$, $\beta_1 = \frac{\alpha(1+\phi)}{1-\alpha} > 0$, $\beta_2 = \beta_3 = \phi > 0$, and $\theta = -\frac{\alpha\gamma}{1-\alpha} < 0$. The error term or region-specific shock, ε_i, is assumed to be identically and independently distributed (iid) for $i = 1, \ldots, N$. Accounting for the interdependence between regions, the equation above

[30] The two different diagonal matrices in Equation (36) are both of dimension $N \times N$, and their general terms are given by $\frac{n_i}{n_i+g^w+\delta_i}$ and $\frac{g^w+\delta_i}{n_i+g^w+\delta_i}$, respectively, with $i = 1, \ldots, N$ (Ertur and Koch, 2011, 250).

[31] For simplicity, the time index has been set to $t = 0$ and is omitted.

can be rewritten in matrix form as

$$\boldsymbol{y} = \boldsymbol{\iota}\beta_0 + \boldsymbol{X\beta} + \theta\boldsymbol{WZ} + \gamma\boldsymbol{Wy} + \boldsymbol{\varepsilon}. \tag{38}$$

This specification is a Spatial Durbin Model (SDM) as it includes spatial lags of the exogenous as well as endogenous variables (LeSage and Pace, 2009).[32] The list below provides an overview of variable definitions in this specification:

- $\boldsymbol{y}$ is an $N \times 1$ vector of the natural logarithm of real income per worker,
- $\boldsymbol{\iota}$ is an $N \times 1$ vector of ones,
- β_0 is a scalar,
- $\boldsymbol{X}$ is an $N \times 3$ matrix of the explanatory variables (the investment rate in physical capital, $s_{K,i}$, divided by the effective depreciation rate, $n_i + g^w + \delta_i$, the growth rate of the number of workers, n_i, and the investment rate in R&D, $s_{A,i}$ – all in logs),
- $\boldsymbol{\beta}$ is a 3×1 vector $[\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)']$ of the regression parameters for the explanatory variables,
- θ is a scalar,
- $\boldsymbol{W}$ is the $N \times N$ interaction matrix (or spatial weight matrix) in non row-normalized form,
- $\boldsymbol{Z}$ is the $N \times 1$ vector of the investment rate in physical capital divided by the effective depreciation rate,
- $\boldsymbol{WZ}$ is the $N \times 1$ vector of the spatial lag of the investment rate in physical capital divided by the effective depreciation rate,
- γ is the spatial autoregressive coefficient,
- $\boldsymbol{Wy}$ is an $N \times 1$ vector denoting the spatial lag of the endogenous variable,
- $\boldsymbol{\varepsilon}$ is an $N \times 1$ vector of errors with mean zero and variance $\sigma^2\boldsymbol{I}$ so that $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ holds.

[32] To be more precise, Equation (38) is a constrained version of the standard Spatial Durbin Model, since in this case only a subset of the potential spatial lags of the exogenous variables is included.

The model specified in Equation (38) nests a series of growth models as special cases of the multi-region Schumpeterian growth model. For instance, the familiar Solow model (see, for example, the original contributions by Solow (1956) and Swan (1956)) is a special case of Equation (37). It results when no interaction (or technological interdependence) between regions exists and consequently $\gamma = 0$ (compare Equation (23)). Furthermore, in the standard Solow model, R&D expenditures are not present, which implies $\phi = 0$.[33] With these conditions, it follows from Equation (37) that in this case steady-state income per worker is given by

$$\ln y_i = \beta_0^S + \beta_1^S \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \varepsilon_i^S. \tag{39}$$

Written in matrix form, this is equivalent to $\boldsymbol{y} = \beta_0 \boldsymbol{\iota} + \beta_1^S \boldsymbol{X^S} + \boldsymbol{\varepsilon^S}$ with $\boldsymbol{X^S}$ an $N \times 1$ vector of the investment rate in physical capital divided by the effective depreciation rate, β_1^S the corresponding regression parameter, and $\boldsymbol{\varepsilon^S}$ an iid vector for the error terms.

Next, the Schumpeterian model by Howitt (2000) and Aghion and Howitt (1998) is also a special case of the multi-region Schumpeterian model as these authors abstain from modeling spillovers due to investment in physical capital (implying $\theta = 0$) and assume that the amount of knowledge that diffuses to other regions is identical for all regions (Howitt, 2000, 838). Hence, if the amount of knowledge diffusion is independent of the specific region, the term $\gamma H_i \sum_{j \neq i}^{N} v_{ij} \ln y_j$ in Equation (37) can be subsumed into the constant of the empirical specification. The result then is

$$\ln y_i = \beta_0^H + \beta_1^H \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \beta_2^H \ln s_{A,i} + \beta_3^H \ln n_i + \varepsilon_i^H \tag{40}$$

which in matrix form reads $\boldsymbol{y} = \beta_0^H \boldsymbol{\iota} + \boldsymbol{X^H} \boldsymbol{\beta^H} + \boldsymbol{\varepsilon^H}$ with $\boldsymbol{X^H}$ an $N \times 3$ matrix of the regressors specified in the equation above and $\boldsymbol{\beta^H}$ the 3×1 vector of corresponding coefficients. The error in this specifications is also iid.

Finally, given that R&D investment has no impact on the Poisson arrival rate and thus $\phi = 0$, it follows that $\beta_2 = 0 = \beta_3$ in Equation (37), and the resulting model is the spatially augmented Solow model developed in Ertur and Koch (2007).[34] Formally, this

[33] There exist extensions of the model, which include this variable. See, for example, Nonneman and Vanhoudt (1996) or Keller and Poutvaara (2005). Additional augmentations of the standard Solow model have been developed, too. These include extending the model by human capital (Mankiw et al., 1992), by health (Knowles and Owen, 1995), by IQ and longevity Ram (2007), or by history (Dalgaard and Strulik, 2013). These models are, however, not nested in the multi-region Schumpeterian model discussed here and hence not estimated.

[34] The model presented in Equation (41) differs from from the one in Ertur and Koch (2007) with respect to the interaction matrix $\boldsymbol{W}$, as in their contribution the matrix of the human capital stock, $\boldsymbol{H}$, is absent.

specification is given by

$$\begin{aligned}\ln y_i = \beta_0^{EK} + \beta_1^{EK} \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \theta^{EK} H_i \sum_{j \neq i}^{N} v_{ij} \ln \frac{s_{K,j}}{n_j + g^w + \delta_j} \\ + \gamma^{EK} H_i \sum_{j \neq i}^{N} v_{ij} \ln y_j + \varepsilon_i^{EK}.\end{aligned} \tag{41}$$

which is a Spatial Durbin Model. In matrix notation, it is given as $\boldsymbol{y} = \beta_0^{EK}\boldsymbol{\iota} + \beta^{EK}\boldsymbol{X^{EK}} + \theta^{EK}\boldsymbol{WX^{EK}} + \gamma^{EK}\boldsymbol{Wy} + \boldsymbol{\varepsilon^{EK}}$ with $\boldsymbol{X}^{EK}$ an $N \times 1$ vector of the values for the investment rate in physical capital divided by the effective depreciation rate, $\boldsymbol{WX^{EK}}$ the spatial lag of this variable, $\boldsymbol{Wy}$ the spatial lag of the dependent variable, and $\boldsymbol{\varepsilon^{EK}}$ the iid error term.

4.2 Estimation Strategy

As LeSage and Pace (2009) point out, a Spatial Durbin Model can be equivalently expressed as a Spatial Autoregressive Model (SAR). Rewriting Equation (38) accordingly, leads to

$$\boldsymbol{y} = \gamma\boldsymbol{Wy} + \tilde{\boldsymbol{X}}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{42}$$

with $\tilde{\boldsymbol{X}} = [\boldsymbol{\iota}\, \boldsymbol{X}\, \boldsymbol{WZ}]$ an $N \times 5$ matrix and $\boldsymbol{\delta} = [\beta_0\, \boldsymbol{\beta}\, \theta]'$ a 5×1 vector. In reduced form, this model is therefore given by[35]

$$\boldsymbol{y} = (\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}\tilde{\boldsymbol{X}}\boldsymbol{\delta} + (\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}.$$

Note that this reduced-form specification implies that the spatial lag of the endogenous variable is correlated with the error term, i.e.

$$Cov[(\boldsymbol{Wy}), \boldsymbol{\varepsilon}] = E[(\boldsymbol{Wy})\boldsymbol{\varepsilon}'] - E[\boldsymbol{Wy}] = \boldsymbol{W}(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}\sigma^2.$$

Hence, ordinary least squares (OLS) estimators will not be consistent.

An alternative to using OLS to estimate the model is provided by Maximum Likelihood (ML) estimation (compare e.g. Lee, 2004). This requires making a distributional assumption for the error terms. Above, it was assumed that the error terms follow a

[35] On the existence of the inverse, see Appendix B.9.

normal distribution, and in this case the log-likelihood function reads

$$\begin{aligned}\ln L(\boldsymbol{y};\boldsymbol{\delta},\gamma,\sigma^2) = &-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln(\sigma^2)+\ln|\boldsymbol{I}-\gamma\boldsymbol{W}| \\ &-\frac{1}{2\sigma^2}\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]'\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right].\end{aligned} \tag{43}$$

In particular, the presence of the determinant $\ln|\boldsymbol{I}-\gamma\boldsymbol{W}|$ in this expression might not be immediately obvious. The following derivation of the function above therefore sheds some light on this term.

4.2.1 Derivation of the Log-likelihood Function

Given the distributional assumption made above for the error (or disturbance) terms, ε_i, in a given region, these have the following probability density function

$$f(\varepsilon_i;0,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right)$$

so that the joint density function of the error terms reads

$$\begin{aligned}f(\varepsilon_1,\ldots,\varepsilon_N;\boldsymbol{0},\sigma^2\boldsymbol{I}) &= \prod_{i=1}^{N} f(\varepsilon_i;0,\sigma^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N\prod_{i=1}^{N}\exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right) \\ f(\boldsymbol{\varepsilon};\boldsymbol{0},\sigma^2\boldsymbol{I}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N\exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right)\end{aligned}$$

where the last line follows from $\sum_{i=1}^{N}\varepsilon_i^2=\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$. However, the disturbance terms cannot be observed, and therefore the likelihood function needs to be based on $\boldsymbol{y}$, which is observable (Anselin, 1988b, 62). Hence, the vector of random variables $\boldsymbol{\varepsilon}$ needs to be transformed into the vector of random variables $\boldsymbol{y}$. This works with the help of a general result on the transformation of variables. It holds that the joint density function $g(\cdot)$ for $\boldsymbol{y}$ is given by (Davidson and MacKinnon, 2004, 430-431)

$$g(\boldsymbol{y})=f(\boldsymbol{\varepsilon})\cdot\left|\frac{\partial\boldsymbol{\varepsilon}}{\partial\boldsymbol{y}}\right|.$$

Due to this result, the determinant will enter the likelihood function. This determinant is also called the Jacobian (determinant) of the transformation (see, for example, Greene, 2003, 844-45). From Equation (42) it follows that the vector of disturbances is given by

$$\boldsymbol{\varepsilon} = (\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}. \tag{44}$$

Therefore, the Jacobian determinant for this case reads $\left|\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{y}}\right| = |\boldsymbol{I} - \gamma \boldsymbol{W}|$. Accordingly, the joint density function for $\boldsymbol{y}$ is

$$g(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \cdot \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) \cdot |\boldsymbol{I} - \gamma \boldsymbol{W}|\,.$$

As the likelihood function coincides with the joint density function (Verbeek, 2004, 164), it can be expressed as

$$L(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) \cdot |\boldsymbol{I} - \gamma \boldsymbol{W}|\,.$$

Inserting for $\boldsymbol{\varepsilon}$ from Equation (44), and taking the natural logarithm of this expression results in

$$\begin{aligned} \ln L(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) = &-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|\boldsymbol{I} - \gamma \boldsymbol{W}| \\ &- \frac{1}{2\sigma^2}\left[(\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]'\left[(\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}\right] \end{aligned}$$

which is identical to Equation (43) above.[36]

Finding the ML estimator, requires maximizing the log-likelihood function with respect to the parameters $\boldsymbol{\delta}, \gamma$, and σ^2, i.e. setting the 5×1 score vector equal to the corresponding zero vector (Verbeek, 2004, 166-167). This multivariate optimization problem can be transformed into a univariate one by concentrating the log-likelihood function with respect to $\boldsymbol{\delta}$ and σ^2. The approach (see Pace and Barry, 1997, 235-236) is to substitute closed-form solutions for the estimators, $\hat{\boldsymbol{\delta}}(\gamma)$ and $\hat{\sigma}^2(\gamma)$, that depend only on the data and the unknown parameter γ, into Equation (43). These solutions can be derived from the first-order conditions for $\boldsymbol{\delta}$ and σ^2 (LeSage and Pace, 2009, 47). The resulting concentrated log-likelihood function can then be maximized with respect to the parameter γ to obtain an estimate, $\hat{\gamma}$, for this parameter. This estimate can in turn be used to back out estimates for the other parameters from the expressions for $\hat{\boldsymbol{\delta}}(\hat{\gamma})$ and $\hat{\sigma}^2(\hat{\gamma})$ (LeSage and Pace, 2009, 47).

[36]The log-likelihood function in the standard regression model might be more familiar, but no determinant occurs in that expression. The reason for the difference is that in the standard case where $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$ the Jacobian is equal to 1. In general, the presence of the determinant in the formula for the transformation ensures that after the transformation the volume under the joint probability density function is still equal to unity (LeSage and Pace, 2009, 80).

4.2.2 Derivation of the Concentrated Log-likelihood Function

Following the approach outlined above, the derivative of the log-likelihood function with respect to σ^2 yields[37]

$$\frac{\partial \ln L(\cdot)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]'\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]. \tag{45}$$

Setting this derivative equal to zero, leads to the maximum likelihood estimator for σ^2, i.e.

$$\hat{\sigma}^2(\gamma) = \frac{1}{N}\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}\right]'\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}\right]. \tag{46}$$

Taking the derivative of Equation (43) with respect to $\boldsymbol{\delta}$ and solving for the maximum likelihood estimator $\hat{\boldsymbol{\delta}}$ is a little more involved so that at this point only the result is presented, while the detailed derivation is delegated to Appendix C.1. The estimator is given by

$$\hat{\boldsymbol{\delta}} = \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}. \tag{47}$$

Defining $\hat{\boldsymbol{\delta}}_{\boldsymbol{O}} \equiv \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}$ and $\hat{\boldsymbol{\delta}}_{\boldsymbol{L}} \equiv \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y}$, the estimator can be equivalently expressed as[38]

$$\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_{\boldsymbol{O}} - \gamma\hat{\boldsymbol{\delta}}_{\boldsymbol{L}}.$$

Defining furthermore the estimated residuals of a regression of $\boldsymbol{y}$ on $\tilde{\boldsymbol{X}}$ as $\hat{\boldsymbol{e}}_{\boldsymbol{O}} \equiv \boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\boldsymbol{O}}$ and the estimated residuals of a regression of $\boldsymbol{W}\boldsymbol{y}$ on $\tilde{\boldsymbol{X}}$ as[39] $\hat{\boldsymbol{e}}_{\boldsymbol{L}} \equiv \boldsymbol{W}\boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\boldsymbol{L}}$, the maximum likelihood estimator $\hat{\sigma}^2$ can be expressed as

$$\hat{\sigma}^2(\gamma) = \left[\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})}{N}\right].$$

Substituting this estimator into the log-likelihood function in Equation (43), yields

$$\begin{aligned}
\ln L(\boldsymbol{y};\gamma) = & -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left[\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})}{N}\right] + \ln|\boldsymbol{I}-\gamma\boldsymbol{W}| \\
& - \frac{1}{2\frac{1}{N}}\cdot\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})}{(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})} \\
= & -\frac{N}{2}\left[\ln(2\pi)+1\right] + \ln|\boldsymbol{I}-\gamma\boldsymbol{W}| - \frac{N}{2}\ln\left[\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}}-\gamma\hat{\boldsymbol{e}}_{\boldsymbol{L}})}{N}\right] \qquad (48)
\end{aligned}$$

[37] Alternatively, taking σ as the parameter in the log-likelihood function instead of σ^2 would lead to identical results in the end.

[38] This is an unbiased estimate. See Keilbach (2000, 153) for the proof.

[39] In Ertur and Koch (2011, 233) there is a slight mistake as they state (converted to the notation used here) that $\hat{\boldsymbol{e}}_{\boldsymbol{L}} = \boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\boldsymbol{L}}$, whereas the expression given in the main text above is the correct one.

which now only depends on the parameter γ. Computation of the maximum likelihood estimator $\hat{\gamma}$ is facilitated by taking recourse to a result by Ord (1975, 121). This result states that the determinant $|\boldsymbol{I}-\gamma\boldsymbol{W}|$ can be expressed in a simpler way via the eigenvalues $\lambda_i, \ldots, \lambda_N$ of the interaction matrix. In particular, it holds that $|\boldsymbol{I}-\gamma\boldsymbol{W}| = \Pi_{i=1}^{N}(1-\gamma\lambda_i)$ or, after taking the natural logarithm: $\ln|\boldsymbol{I}-\gamma\boldsymbol{W}| = \sum_{i=1}^{N}\ln(1-\gamma\lambda_i)$. This latter result is substituted into the log-likelihood in Equation (48). The advantage of employing this expression is that in the numerical optimization procedure for the determination of $\hat{\gamma}$, the eigenvalues need only be determined once (Ertur and Koch, 2011, 233). Having determined $\hat{\gamma}$ numerically, the value can be substituted into the closed-form solutions for $\hat{\sigma}^2$ and $\hat{\boldsymbol{\delta}}$ in Equations (46) and (47) to obtain the estimates for these parameters.

5 Data, Estimation Results, and Interpretation of Model Parameters

This section first provides a detailed overview of the data and the construction of the variables for the empirical analysis. Afterwards estimation results of the models specified in Section 4.1 will be presented and discussed. Estimates for the direct, indirect and total impacts of the variables in the spatial models will also be presented.

5.1 Data

The empirical analysis focusses on the US federal states. As is common practice in studies analyzing US economic development on a state level, Alaska, Hawaii, and (by definition) Washington, D.C. are dropped from the sample so that only the 48 contiguous (or continental) states are included.[40] In addition, following the approach by Bode et al. (2012, 27), Delaware is also excluded so that the baseline sample consists of 47 states. The state of Delaware is home to a large financial industry, and it might be the case that this characteristic influences the estimation results. Also, as Hanushek et al. (2015, 16) note, gross state product (GSP) in Delaware might not be well described by a standard production function, as more than 35% of its GSP in 2007 is accounted for by finance and industry, whereas the remaining states only reach less than half this value.[41]

[40] Compare, for instance, Holtz-Eakin (1993), Barro and Sala-i-Martin (1992) or, more recently, Yamarik (2011) for this composition of the sample.

[41] Hanushek et al. (2015, 16) quote figures from an article in The Economist (2013) stating that Delaware is a tax haven where companies outnumber people (945,000 vs. 917,092).

The sample period in the empirical analysis below covers the 11 years from 1997-2007. This period is rather short, but still in line with the studies (on different units of observation) by, for instance, Ertur and Koch (2011, 235), who analyze a period of 14 years or Fischer (2011, 430) and Fischer et al. (2010, 592), who have data for 10 years.[42] For the present analysis, data for more recent years is available in the case of a subset of the variables used in the analysis. The reason 2007 is chosen as the final year is twofold: On the one hand, it is chosen to avoid the financial crisis starting in 2008 influencing the results, and, on the other hand, data for the investment in physical capital is only available up to 2007. For years prior to 1997 data is available for many variables pertaining to the analysis. However, 1997 is chosen as a cutoff, since the time series for the dependent variable has a structural break in that year.[43]

Output, y_i, is measured as real chained-weighted gross state product generated in the private sector measured in 2000 dollars, and the data stems from the Bureau of Economic Analysis' (BEA) regional accounts data (BEA, 2015b). The variable is constructed by dividing nominal gross state product generated in the private sector by the implicit price deflators for the gross domestic product (GDP), which is taken from the national accounts data of the BEA (2015a).[44] In more detail, the following approach is employed (compare Peri, 2012, 350): The time series for the GDP deflator is from the BEA (2015a) and has 2009 as its reference year. Therefore, the reference year for this series is first changed to the year 2000 before using these values to convert GSP in nominal dollars to GSP in 2000 real dollars.[45]

Labor is measured as total employment on private payrolls, as in, for example, Yamarik (2013). This data is reported by the Bureau of Labor Statistics (2015) in its Current Employment Statistics, and n_i is the average annual growth rate of total employment.[46]

[42]As Fischer (2011, 429) notes with reference to Durlauf and Quah (1999) and Islam (1995), steady-state regressions are valid for relatively short time periods.

[43]The break occurs as the US shifted from the Standard Industrial Classification (SIC) to the North American Industry Classification System (NAICS) and the Bureau of Economic Analysis on their website strongly "cautions against appending the two data series" (compare `http://www.bea.gov/regional/docs/product/` (accessed: 11 August, 2015)).

[44]This is similar to, for instance, Yamarik (2006) and Barro and Sala-i-Martin (2004), who use the national consumer price index to deflate nominal personal income.

[45]As Barro and Sala-i-Martin (2004, 497) note: "As long as the same index is used at each date for each state, the particular index chosen does not affect the relative levels and growth rates across states".

[46]The values for this variable pose a slight problem for the estimation in the next section. The model is specified in logs, but, as the summary statistics in Table 1 show, the minimum value for the employment growth rate is -0.5%, for which the logarithmic transformation is not defined. Besides this value for Michigan, also Ohio has a negative employment growth rate over the period (-0.01%). Several approaches exist to deal with this issue. The one preferred here for its simplicity, follows Sarel (1996, 203), who encounters this problem in the context of inflation rates. He sets the negative values equal to the smallest positive observed value in the sample. For the present analysis, the respective value is 0.23% (for Mississippi). Alternatively, for comparable situations, it is suggested to add a constant

Values for the state-level real investment rate, $s_{K,i}$, are not available from official US agencies. However, Yamarik (2013), updating a previous contribution by Garofalo and Yamarik (2002), provides values for state-level real investment in 2000 dollars. Dividing those by the real GSP values then leads to the values for the state-level real investment rate. Furthermore, in Yamarik (2013), annual values for state-specific depreciation rates of physical capital, δ_i, are also provided so that here, in contrast to other studies, it is possible to deviate from the assumption of identical depreciation rates across all units of observation and use the average state-specific annual depreciation rate of physical capital in the empirical study instead.[47] The growth rate, g_w, is set to 0.02, which is in line with the value chosen by Howitt (2000, 841) and also is similar to the approach by Yamarik (2006) considering that he obtains a mean value of 9% for $n_i + g_w + \delta_i$ where the sample covers the time period 1950-2000.[48] Investment in R&D, $s_{A,i}$, is measured as the average real research and development expenditure as a percentage of real gross state product. Data for this variable is provided by the Organisation for Economic Co-operation and Development's (OECD) Regional Database (OECD, 2015). Total R&D expenditures are given by summing up expenditures in the business, government, higher education, and private non-profit sectors (OECD, 2015). Concerning the human capital stock, H_i, this variable is measured by the average share of individuals above the age of 24 with four or more years of college (more specifically, a Bachelor's degree or higher). This is in accordance with the measure used by, for example, Bode et al. (2012) or Yamarik (2006). The data is supplied by the Current Population Survey of the United States Census Bureau (2015). For this variable, no state-level data is available for 2007 so that this year is omitted in calculating the average values.[49]

With respect to the interaction matrix $\boldsymbol{W}$, it is important to highlight that the weights should be exogenous to the variables in the model (Ertur and Koch, 2007, 1042). This restricts the choice of variables that might be considered to model connectivity between states considerably. In general, studies have relied on geographic distance to specify the weights in the interaction (or spatial weight) matrix. This measure allowed researchers to

to the variable before applying the logarithmic transformation to ensure that all values are positive (see, for example, Dowdy et al. (2004, 329) or Wooldridge (2013, 193)). Adding, for instance, 0.006 to all employment growth rates before taking logs ensures that the growth rate for Michigan is positive but small (0.1%). Thirdly, the observations for Michigan and Ohio could be dropped from the sample. All three methods of handling this problem lead to only minor quantitative changes in the estimated parameters. Note that the adjustment is only necessary for the employment growth rate variable, but not for the effective depreciation rate, $n_i + g^w + \delta_i$, which is positive for all observations.

[47] The data for $s_{K,i}$ and δ_i is available on Steven Yamarik's website under: `https://web.csulb.edu/~syamarik/` (accessed: 11 August, 2015).

[48] Assuming $g_w = 0.02$, the average value of $n_i + g_w + \delta_i$ in the present sample is approximately 8% (see Table 1).

[49] See the user note at the following link: `http://www.census.gov/hhes/socdemo/education/data/cps/2007/usernote.html` (accessed: 11 August, 2015).

capture that effects between units of observations diminish with geographic distance (see, for example, Eaton and Kortum (1996) or Keller (2002)).[50] The distance-decay effect can be formalized in a variety of ways. Here, three different interaction matrices of the form $\boldsymbol{W} = \boldsymbol{HV}$ with general weights given by $w_{ij} = H_i v_{ij}$ will be considered to assess the robustness of the empirical results. As will be clear from the functional forms specified in Equations (49), (50), and (51) below, the matrices $\boldsymbol{V_1}, \boldsymbol{V_2}$ and $\boldsymbol{V_3}$ are row standardized, whereas the matrices $\boldsymbol{W_1}, \boldsymbol{W_2}$ and $\boldsymbol{W_3}$ are not, as they are multiplied by the matrix $\boldsymbol{H}$.

The first interaction matrix, $\boldsymbol{W_1}$, is based on a binary first-order contiguity matrix as in Fischer (2011, 430) or Rey and Montouri (1999, 146). States are considered contiguous (or, more simply, neighbors), if they share a common border (i.e. Montana and North Dakota) and the modifier "first-order" refers to the fact that only direct neighbors are relevant[51] so that Minnesota is a first-order neighbor of North Dakota, but a second-order neighbor of Montana (see the map in Figure 1).[52] In formal terms, the weights in matrix $\boldsymbol{W_1}$ are therefore described by $w_{ij}(1) = H_i v_{ij}(1)$, with

$$v_{ij}(1) = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{\sum_{j \neq i}^{N} v_{ij}(1)} & \text{if } i \text{ and } j \text{ are neighbors.} \end{cases} \tag{49}$$

A second possibility to model the distance-decay effect abstracts from the binary option adopted above and connects all states directly with each other. The weights in matrix $\boldsymbol{W_2}$ are given by $w_{ij}(2) = H_i v_{ij}(2)$, and, as, for example, in Ertur and Koch (2011), the following continuous functional form is assumed for these weights

$$v_{ij}(2) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-d_{ij}}}{\sum_{j \neq i}^{N} e^{-d_{ij}}} & \text{otherwise.} \end{cases} \tag{50}$$

Here, d_{ij} is the great circle distance – the shortest path between two points on the surface of a sphere – between the geographic centroids of the US states. These centroids

[50]Measures based on geographic distance are, however, not the only possibility. Another exogenous measure is genetic distance between units of observation. In the present analysis, it is unfortunately not possible to use this alternative measure, since the necessary bilateral distance measures are not available on the level of US states (see Spolaore and Wacziarg (2009) for the relevant country-level data). A similar issue arises for another potential candidate measure, linguistic distance, that has been used in studies applying spatial econometric methods (compare, for example, Isphording and Otten (2013) or Melitz and Toubal (2014)).

[51]One might wonder about the quadripoint where the borders of Colorado, New Mexico, Arizona and Utah meet (see Figure 1). In the present analysis, the pairs Arizona/Colorado and Utah/New Mexico are considered neighbors.

[52]Note that this specification does not rule out spillovers from Minnesota to Montana, as all states are connected via the spatial multiplier (compare Equation (32)).

are illustrated by the black dots in Figure 4.

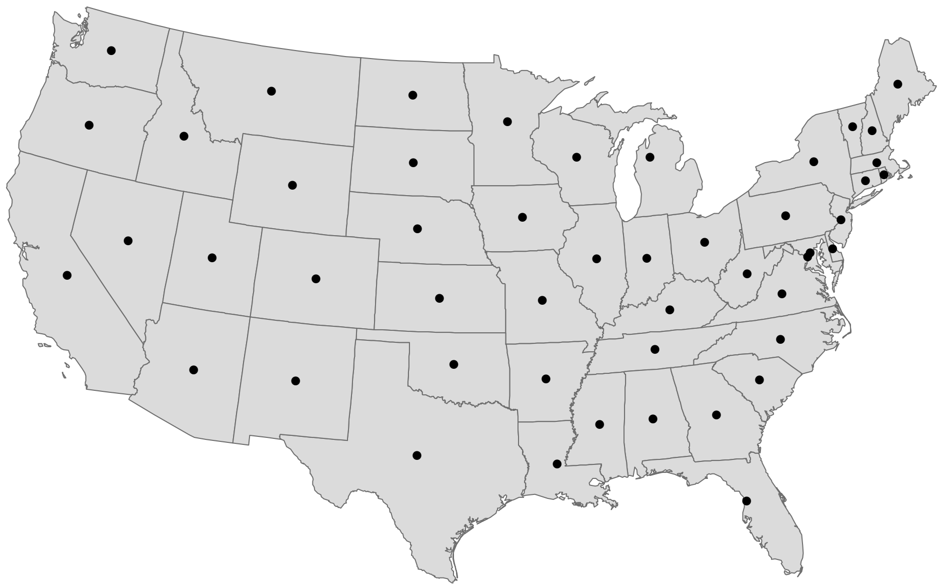


Figure 4: Geographic Centroids of US States.

The third interaction matrix, $\boldsymbol{W_3}$, has weights $w_{ij}(3) = H_i v_{ij}(3)$ and has a similar form to the matrix in, for example, Bode et al. (2012) and Basile (2014). It adopts the negative exponential form of matrix $\boldsymbol{W_2}$, but scales it with a factor τ. In addition, a distance cutoff is introduced. If the distance between the centroids of the two states is larger than this threshold, the corresponding matrix entry is set to zero, implying that direct spillovers between these states are non-existent. Formally, the matrix entries are calculated by

$$v_{ij}(3) = \begin{cases} 0 & \text{if } i = j \text{ or if } d_{ij} > 512\text{km} \\ \frac{e^{-\tau d_{ij}}}{\sum_{j \neq i}^{N} e^{-\tau d_{ij}}} & \text{if } d_{ij} < 512\text{km}. \end{cases} \tag{51}$$

As in Bode et al. (2012), τ is set to 0.02, and the distance cutoff is chosen following Basile (2014, 12) as the minimum distance ensuring that all states have at least one neighbor. For the present sample, this distance is slightly below 512km (the distance between the centroids of Arizona and New Mexico). Concerning τ, Bode et al. (2012) supply a helpful illustration: They argue that the weights in the interaction matrices can be understood similar to iceberg transportation costs[53] where the parameter τ indicates the percentage of knowledge diffusion that is lost per kilometer. For $\tau = 0.02$ this implies that after 50

[53]These are familiar from new economic geography (see, e.g. Krugman, 1991, 489). Samuelson described the general concept in the following way: "To carry [a] good across the ocean you must pay some of the good itself" and illustrated it more specifically by continuing that "only a fraction of ice exported

kilometers $1 - e^{-0.02*50\text{km}} \approx 63.2\%$ of the iceberg "has melted away" and approximately 86.5% after 100 kilometers.

Before presenting the estimation results in the following section, Table 1 provides summary statistics for the variables used in the empirical analyses.

Table 1: Summary Statistics – Baseline Sample.

Variable	Mean	Median	Standard deviation	Minimum	Maximum
y_i	85,012.07	80,896.36	13,921.60	66,616.49	123,281.63
$s_{K,i}$	0.085	0.081	0.017	0.063	0.139
n_i	0.012	0.011	0.008	-0.005	0.037
δ_i	0.048	0.048	0.002	0.044	0.051
$n_i + g^w + \delta_i$	0.080	0.079	0.008	0.062	0.101
$s_{A,i}$	0.022	0.019	0.015	0.005	0.075
H_i	0.255	0.244	0.046	0.158	0.351
$\frac{s_{K,i}}{n_i+g^w+\delta_i}$	1.064	1.034	0.172	0.862	1.706
$\boldsymbol{W_1 s_K}$	0.264	0.261	0.051	0.159	0.416
$\boldsymbol{W_2 s_K}$	0.274	0.262	0.070	0.162	0.494
$\boldsymbol{W_3 s_K}$	0.271	0.260	0.061	0.159	0.438
$\boldsymbol{W_1 y}$	21,073.61	20,214.21	4,628.53	12,977.69	32,609.39
$\boldsymbol{W_2 y}$	21,302.32	20,192.91	5,446.95	12,398.25	42,844.25
$\boldsymbol{W_3 y}$	21,616.61	19,961.00	5,454.27	13,399.71	38,037.59

Note: The given values are the original values (i.e. not in logs) for the benchmark sample of 47 states and the period 1997-2007 with y_i the income per worker in 2007.

5.2 Estimation Results

Table 2 shows the estimation results[54] for the series of models described in Section 4.1. In Column 1, the standard Solow model from Equation (39) is estimated by ordinary least squares (OLS), and the results show that, in line with the predictions of this model, the investment rate in physical capital divided by the effective depreciation rate has a positive and significant impact on steady-state income per worker (p-value = 0.033). As the model

reaches its destination as unmelted ice" (Samuelson, 1954, 268). However, the general idea goes back almost two centuries to von Thünen, who noted with respect to the transport of grain by horse-drawn carriage that if the distance between farm and city (and back to the farm) is large enough (50 miles in the specific example he describes), then "ist also der Transport des Korns auf 50 Meilen unmöglich, weil die ganze Ladung oder deren Werth auf der Hin- und Zurückreise von den Pferden und den dabei angestellten Menschen verzehrt wird" (von Thünen, 1826, 9).

[54]All estimations have been conducted in Matlab with the Spatial Econometrics Toolbox provided by LeSage. The toolbox is available under: `http://www.spatial-econometrics.com/` (accessed: 11 August, 2015).

is specified in logs, the estimated coefficient points to an increase of approximately 3.3% due to a 10% increase in the investment rate in physical capital.

Table 2: Estimation Results for Three Different Models for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Model	Solow (1956)	Howitt (2000)	Ertur and Koch (2007)			Ertur and Koch (2011)		
Interaction matrix			$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Constant	11.322	11.690	11.019	10.966	10.965	11.253	10.977	10.977
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.326	0.362	0.310	0.325	0.282	0.329	0.321	0.278
	(0.033)	(0.017)	(0.023)	(0.023)	(0.053)	(0.018)	(0.029)	(0.064)
$\ln s_{A,i}$	—	0.065	—	—	—	0.023	−0.006	−0.005
		(0.041)				(0.494)	(0.868)	(0.906)
$\ln n_i$	—	0.023	—	—	—	0.018	0.010	0.008
		(0.467)				(0.523)	(0.717)	(0.769)
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	—	—	−1.728	−0.567	−0.237	−1.641	−0.606	−0.268
			(0.096)	(0.263)	(0.719)	(0.118)	(0.239)	(0.687)
γ	—	—	0.117	0.126	0.126	0.096	0.131	0.129
			(0.004)	(0.000)	(0.001)	(0.048)	(0.009)	(0.011)
AIC	−3.794	−3.813	−3.896	−3.965	−3.942	−3.829	−3.885	−3.860
BIC	−3.715	−3.655	−3.738	−3.807	−3.785	−3.593	−3.648	−3.624
Number of observations	47	47	47	47	47	47	47	47

Note: p-values are given in parentheses.

Column 2 shows the estimation results from the Howitt model, specified in Equation (40). The coefficient for the investment rate over the effective depreciation rate has increased slightly, quantitatively as well as in significance, compared to the estimation of the Solow model. The newly added variable, investment in R&D, is estimated to have a positive and significant effect (p-value $= 0.041$) on per worker income in steady state. However, the effect is smaller than for investments in physical capital, as a 10% increase in R&D investment would result in a 0.65% increase in per worker income in steady state. Regarding the remaining variable, the employment growth rate, its effect is not significant with a p-value of 0.467.[55]

The next 3 columns estimate the spatially-augmented Solow model from Equation (41) by maximum likelihood as described in Section 4.2. Here, the approach differs slightly from Ertur and Koch (2011), as the approach from Basile (2014) is followed to estimate the Spatial Durbin Model instead of the Spatial Error Model (SEM) to obtain estimates for the coefficient of the spatial lag of the investment variable as well.[56] For all three interaction matrices the effect of the investment variable accords with implications derived from

[55]The non-significance of this variable in the Howitt specification is also found by Ertur and Koch (2011) in their cross-country sample.

[56]In contrast to the SDM model, the SEM model contains spatial autocorrelation only in the error

the theoretic model, and, with the exception of the matrix $\boldsymbol{W_3}$, is also significant at the 5%-level. The estimated coefficients for the spatial lag of the investment variable are not significant in either case. Note, however, that the estimate for the spatial autoregressive coefficient is highly significant (at the 1%-level) for all three matrices.

As this model, by definition, contains interaction between regions, and an interdependent system is estimated, a direct interpretation of the estimated parameters as in the case of the models estimated by OLS is not feasible and might lead to invalid conclusions. The next section presents a method developed by LeSage and Pace (2009) to disentangle the direct and indirect impacts in spatial models.

In the remaining three columns, the multi-region Schumpeterian growth model from Equation (37) is estimated. Similar to the case of the spatially-augmented Solow model, the impact of the investment rate in physical capital is positive for all three interaction matrices and significant at the 5%-level for matrices $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$. Also, the spatial lag of this variable is significant in neither case at standard significance levels. Concerning the newly added variables, the investment rate in R&D and the employment growth rate, the estimated coefficients are not significant in either case. However, again the estimate for the spatial autoregressive coefficient, γ, is estimated to be positive and significant for all three matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$, implying that the states cannot be treated as independent observations.

These estimation results do not provide a clear picture, as, for example, the impact of the R&D investment rate, seems to affect income per worker in steady state in the non-spatial model, but not in the spatial model, although the information criteria point to the latter one.[57] Nonetheless, due to the estimates for the parameter γ, it emerges from these results that interaction effects between observations seem to be relevant.[58]

term, but not in the regressors. See Ertur and Koch (2011, 234 and 240-241) for the specific model addressed here and, for instance, LeSage and Pace (2009) on the spatial error model in general.

[57] Akaike's Information Criterion (AIC) and the Schwarz or Bayesian Information Criterion (BIC) are calculated according to the formulae (see, for example, Greene (2003, 160)):

$$\text{AIC} = \log\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}\right) + 2\frac{K}{N} \qquad \text{and} \qquad \text{BIC} = \log\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}\right) + \frac{K}{N}\log N,$$

where $\hat{\varepsilon}$ denotes the residuals of the estimation and K signifies the number of parameters (for the original contributions regarding these information criteria, see Akaike (1973) and Schwarz (1978)). It should be kept in mind here that only nested models can be compared according to these criteria. Accordingly, comparisons are possible across model with the same interaction matrix, but not between, for example, the models in the last two columns.

[58] The goal here is to test empirically the four different types of models that are contained in a "completely integrated theoretical and empirical framework" (Ertur and Koch, 2011, 216). Hence, the subject of model comparison as traditionally understood, is assigned a reduced role here. In the context of comparison of (spatial) econometric models, the two ends of the spectrum are the specific-to-general approach and the general-to-specific approach (see, for example, Le Gallo (2014, 1528-1529) on these

In Section 5.1 it has been mentioned that the state of Delaware has been excluded from the baseline sample due to the presence of a large financial and insurance sector. Appendix D.1 presents the estimation results when Delaware is included in the sample (see Table D.1). As it turns out, including the state in the sample, results in the coefficient on the physical investment rate over the effective depreciation rate losing its significance, thereby lending credence to the conjecture that this state might not be well described by the model considered here.[59]

5.3 Interpretation of the Model Parameters

Due to the interaction effects contained in the spatial models via the inclusion of the spatial lags, the coefficient estimates in Table 2 cannot be interpreted directly. At this point, it is helpful to refer back to the elasticities calculated in Equations (31), (35), and (36). These $N \times N$ matrices describe the effects of changes in the explanatory variables on the dependent variables. The individual entries in these matrices denote, for instance, the effect of an increase in the investment rate in R&D in Maine on the per worker income in steady state in North Dakota. It becomes clear that the effects will differ depending on the pairs of states chosen and thus reporting all individual effects is rather unwieldy. LeSage and Pace (2009) helpfully provide a method to summarize in a clear manner the estimation results on the direct and indirect effects (or spillovers).[60] The direct effects are the partial derivatives measuring the change in the dependent variable in region i due to a change in the explanatory variable in region i. These effects are measured on the diagonal of the matrix of elasticities (compare Equation (33)). LeSage and Pace (2009) suggest to summarize the direct impact with the average value of the diagonal matrix elements.

A change in the explanatory variable in region i also affects the dependent variable in the other regions, and these indirect impacts are captured by the off-diagonal entries in the matrix (compare Equation (34)). With regard to this effect, the proposed summary measure is the average of the row sum of these off-diagonal matrix entries. This row

approaches). The former strategy has been found to outperform the latter strategy in a specific context not including the SDM as a possible specification (Florax et al., 2003). On the other hand, LeSage and Pace (2009) suggest to start with the SDM model, whereas an approach outlined by Elhorst can be seen as a combination of the two search strategies that chooses as a starting point, however, the specific model. A further reason these approaches have not been adhered to strictly here is that they rely on tests which have been specified for row-standardized interaction matrices (compare, for example, Anselin (1988a) and Anselin et al. (1996)). It is not clear, if these can be applied in the given context in a straightforward manner for models in which the interaction matrix is not described by this characteristic.

[59] Also, Washington D.C. has been omitted from the sample. Including it does not lead to qualitative changes in the results compared to the baseline estimates. Detailed estimation results are omitted though.

[60] A very lucid exposition of their approach can be found in Section 6 of Elhorst (2010).

sum measures the effect on the dependent variable in region i due to a change in the explanatory variables in the remaining regions. Straightforwardly, the average of these row sums is then chosen as the summary measure for the indirect effects.[61] Summing up the direct and indirect effects (i.e. all the elements in a row) gives then a measure for the total impact. The average of these sums is chosen as the corresponding summary measure.

Table 3: Estimation Results for the Direct, Indirect and Total Impacts in the Multi-Region Schumpeterian Model for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Interaction matrix	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Direct impacts:			
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.330	0.321	0.278
	(0.022)	(0.034)	(0.070)
$\ln s_{A,i}$	0.023	−0.006	−0.005
	(0.497)	(0.867)	(0.905)
$\ln n_i$	0.018	0.010	0.008
	(0.525)	(0.718)	(0.772)
Indirect impacts:			
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	0.034	0.047	0.039
	(0.149)	(0.105)	(0.150)
$\boldsymbol{W} \ln s_{A,j}$	0.001	−0.003	−0.002
	(0.724)	(0.699)	(0.724)
$\boldsymbol{W} \ln n_j$	0.002	0.001	0.001
	(0.623)	(0.797)	(0.851)
Total impacts:			
$\frac{\ln s_{K,i}}{\ln(n_i+0.02+\delta_i)} + \boldsymbol{W} \frac{\ln s_{K,j}}{\ln(n_j+0.02+\delta_j)}$	0.364	0.368	0.318
	(0.020)	(0.031)	(0.066)
$\ln s_{A,i} + \boldsymbol{W} \ln s_{A,j}$	0.025	−0.009	−0.007
	(0.509)	(0.837)	(0.875)
$\ln n_i + \boldsymbol{W} \ln n_j$	0.020	0.012	0.009
	(0.528)	(0.726)	(0.780)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

Table 3 presents the estimates for the direct, indirect, and total impacts for the multi-region Schumpeterian model calculated in the way just described. The results show that

[61] As LeSage and Pace (2009, 37) demonstrate, an identical value for the indirect effect is obtained by summing up the off-diagonal column elements and calculating the average of these sums. The interpretation is however different, as, for instance, the latter measure captures the impact of a change in the exogenous variable in region i on the dependent variable in all other regions. In the context, of the present model this measure reports, for example, the impact of an increase in R&D investment in Massachusetts on the per worker income in the remaining US states, whereas the sum of the off-diagonal row elements would report the change in, for example, the per worker income in Massachusetts due to a change in the R&D investment rate by an identical amount in the remaining states.

the indirect effects are not significant for any of the variables included in the regression, independently of the specific interaction matrix.[62] Concerning the estimates for the direct and total impacts, these are positive and significant for the investment rate over the effective depreciation rate in the case of interaction matrices $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$, but not if matrix $\boldsymbol{W_3}$ that contains a distance cutoff is included. Quantitatively, the significant estimates point to an increase of approximately 3.6% in per worker income due to a 10% increase in investment in physical capital.[63]

Regarding the sample that includes Delaware, estimates for the impacts are given in Table D.2 in Appendix D.1. They show that, in contrast to the baseline sample, the direct and total impacts are not significant no matter the interaction matrix included.

Summarizing the empirical results with respect to the multi-region Schumpeterian growth model, it needs to be stated that even though the model's implications are borne out for a particular sample in a cross-country analysis in Ertur and Koch (2011), these results are not readily transferable to the sample of US states analyzed here. Whereas R&D investments have a positive impact on income per worker in the Howitt model, in which the amount of knowledge that diffuses between regions is identical (see Ertur and Koch (2011, 238)), this is not the case in its version with more complex spatial interactions. It might be that the inclusion of the spatial lags in the SDM model is not warranted. As Greene (2003, 151) notes, in such a situation the estimates become less precise and therefore are less likely to be significant. Indeed, the results from testing for the presence of spatial autocorrelation in the residuals of the Howitt model with Moran's I test[64] do not point to estimating a spatial version of the model. However, as the results in Table 2 show, the estimate for the spatial autoregressive coefficient is highly significant.[65] The estimation of this model therefore provides new information; in particular, when compared to the results by Basile (2014) for 248 European NUTS 2 regions. He estimates a growth

[62]Inference on the statistical significance of the parameters is based on p-values which have been obtained from simulating the distribution of the respective effects with the help of the variance-covariance matrix derived in Appendix C.2.

[63]For the spatial Solow model, i.e. the SDM model in Columns 3-5 in Table 2, the respective impacts are not significant in any case and detailed results are omitted here.

[64]The test statistic is given by

$$I = \frac{N}{S_0}\left(\frac{\hat{\boldsymbol{\varepsilon}}'\boldsymbol{W}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}\right)$$

where $\hat{\boldsymbol{\varepsilon}}$ are the residuals from the OLS regression and S_0 is a standardization factor, that equals 1 in the case of a row-standardized interaction matrix, as it is given by the sum of all the elements in $\boldsymbol{W}$ (see, for example, Le Gallo (2014, 1524), who also provides the expressions for the expectation and variance of I, derived by Cliff and Ord (1972) under the null hypothesis).

[65]Tentative evidence from more specific Lagrange Multiplier tests, which in contrast to Moran's I test, specify a particular alternative hypothesis also point to including a spatial lag in the Howitt model. Results of these tests are omitted here though due to the possible issues regarding non-standardized interaction matrices mentioned at the end of Footnote 58.

version of the multi-region Schumpeterian model for these regions for the period 1991-2011 and finds that the estimates have the signs implied by theory and are significant.[66] For the US states, it might be that spatial interaction between states exists, which is, however, only captured by the variable income per worker and not by, for instance, R&D investments. The Moran scatterplot in Figure 2 only hinted at potential spillovers from R&D investments, but the econometric analysis finds no support for these.

Before concluding, a final series of estimation results for the baseline sample of observations will be briefly discussed. Despite the warning against appending the two data series for the GSP variable mentioned in Section 5.1, Appendix D.2 ignores this. The results are qualitatively similar to the ones for the shorter sample. However, notable differences in the significance of variables exist, for example, in the Howitt model where the R&D investment rate no longer has a statistically significant impact. In the multi-region Schumpeterian model when matrix $\boldsymbol{W_1}$ is used, the spatial autoregressive coefficient is not significant (compare Table D.4 for these results). A further difference concerns the direct and total impact estimates for the variable investment rate in physical capital over the effective depreciation rate. These have increased in size to values larger than 0.5 and are highly significant with p-values below 0.003 (see Table D.5).[67]

6 Conclusion

In this paper, the multi-region Schumpeterian growth model developed by Ertur and Koch (2011) has been presented in detail. A characteristic feature of this model is that regions are not considered to develop in isolation from each other, but rather interdependence between regions via knowledge spillovers is explicitly included. Technological progress in the model results from purposeful investments in R&D. It has been shown that how much a region can benefit from a given amount of knowledge spillovers depends on the way a region is connected to other regions and on the distance to its own technological frontier.

Also, the econometric strategy to estimate the equation for the steady-state income per worker that results from the theoretical model has been thoroughly outlined. In contrast to the original contribution, the level of aggregation in the empirical analysis has been reduced, and the model's implications have been tested for a sample consisting of states within a single country (the United States) instead of across countries.

[66] He also provides estimates for the direct, indirect, and total impacts, but no information on their significance is given.

[67] Also, in the case of interaction matrix $\boldsymbol{W_2}$, the indirect effect is marginally significant now at the 5%-level.

The estimation results presented here do not provide full support for all implications derived from the theoretic model. For instance, the hypothesis of technological interdependence between the regions receives support, as the parameter gauging this characteristic is estimated to be positive and statistically significant for all three interaction matrices considered. However, a statistically significant impact of, for example, R&D investment on per worker income could not be detected in this model, even though it was present in the model with a simplified interaction structure (i.e. the Howitt (2000) model). Despite this result, the more nuanced way interaction is modeled in Ertur and Koch (2011) may seem more plausible, as these authors assume that the net effect of the knowledge spillovers depends on the absorptive capacity, i.e. the level of human capital in the receiving region, which is in contrast to the more basic assumption that the amount of knowledge diffused by each region to the other regions is identical.

This distinction may also point to an explanation for the differing estimation results. As the OECD notes in its Science, Technology and Industry Outlook: "US firms are at or near the forefront of technological advances in a number of areas" (OECD, 2010, 232), and the "United States has long been, and still is, at the forefront of cutting-edge science, technology and innovation" (OECD, 2014, 444). Moreover, the various US states are heterogeneous. Hence, it might be the case that potential knowledge spillovers from investment in R&D and physical capital arise in the form of highly-specialized knowledge in a given state, and this knowledge might only diffuse to a very low extent, as it cannot be productively used in the states the originating state is connected to. The receiving states might lack the absorptive capacity to benefit from inter-industry spillovers.

From a different perspective, the model does not differentiate between, for example, various types of workers and an identical level of human capital in two states might hide a large diversity in the composition of human capital. If one assumes identical human capital levels in two states that have different industry structures whose requirements are mirrored in the diversity of the respective state's human capital, then the model's mechanics would imply that spillovers originating in the state with a strong presence in e.g. nanotechnology to the state with a large presence in, for example, car manufacturing would necessarily be reflected in an increase in per worker income. However, the knowledge generated in nanotechnology might not be readily applicable in the car manufacturing sector, since human capital in this sector lacks the necessary complementarity to benefit from the knowledge spillovers. The implied impact on per worker income might then not show up in the data.

This paper has deliberately chosen to stay in a similar framework as Ertur and Koch (2011), both theoretically as well as econometrically, to obtain results that are comparable to a certain extent. Naturally, other estimation approaches exist, and future research will focus on estimating, for instance, a spatial panel model for this sample. Also, the specific choice of the interaction matrix is an interesting topic for further study. In the present analysis, even though the estimation results for the three interaction matrices were similar, they were not identical. The method of Bayesian Model Averaging may be a fruitful avenue for finding a matrix that fits the data more closely. Furthermore, as knowledge spillovers decrease with distance, conducting the analysis at e.g. the level of the county or metropolitan area might lead to additional insights. However, data availability is the restricting factor in this case.

Appendix

A Poisson Processes

In the literature on Schumpeterian growth models it is standard to model the occurrence of an innovation via a Poisson arrival rate or to read about Poisson processes (compare, for instance, Aghion and Howitt (1992) and Aghion et al. (2014)).[68] A detailed exposition of these notions is however seldom provided so that these concepts from statistical theory may pose some difficulties at first glance. This appendix therefore serves as a brief review of the general concepts concerning Poisson processes.

An important step towards understanding Poisson processes concerns the exponential distribution. If a continuous random variable X is exponentially distributed, then its probability density function (PDF) is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{A.1}$$

Here, λ is a parameter for which $\lambda > 0$, and x is a particular value of the random variable. This function is illustrated for two different values of λ in Figure A.1.

The cumulative distribution function, which gives the probability that $X \leq x$, with $x \geq 0$ is then given by

$$P\{X \leq x\} = F(x) = \int_0^x f(\tau)\, d\tau = \int_0^x \lambda e^{-\lambda\tau} d\tau = \Big|_0^x - e^{-\lambda\tau} = 1 - e^{-\lambda x}. \tag{A.2}$$

The exponential distribution has the property of being memoryless (Ross, 2010, 294). This property can formally be stated as

$$P\{X > s + t | X > t\} = P\{X > s\} \qquad \forall\, s, t \geq 0.$$

Interpreting the random variable X, for instance, as the lifetime of a certain machine, instrument, or device like a traffic light, the equation above states that the probability that a traffic light functions $s + t$ units of time (i.e. days), given that it already has worked for t days is the same as the unconditional probability that it works for s days. Put differently, the traffic light "does not remember" it has already worked for t days. A concept, known as the hazard rate or failure rate, helps to illustrate this property. For

[68] Poisson processes also play an important role in other areas of economics like labor or monetary economics. Wälde (2011, 261) provides a brief list of applications in economics.

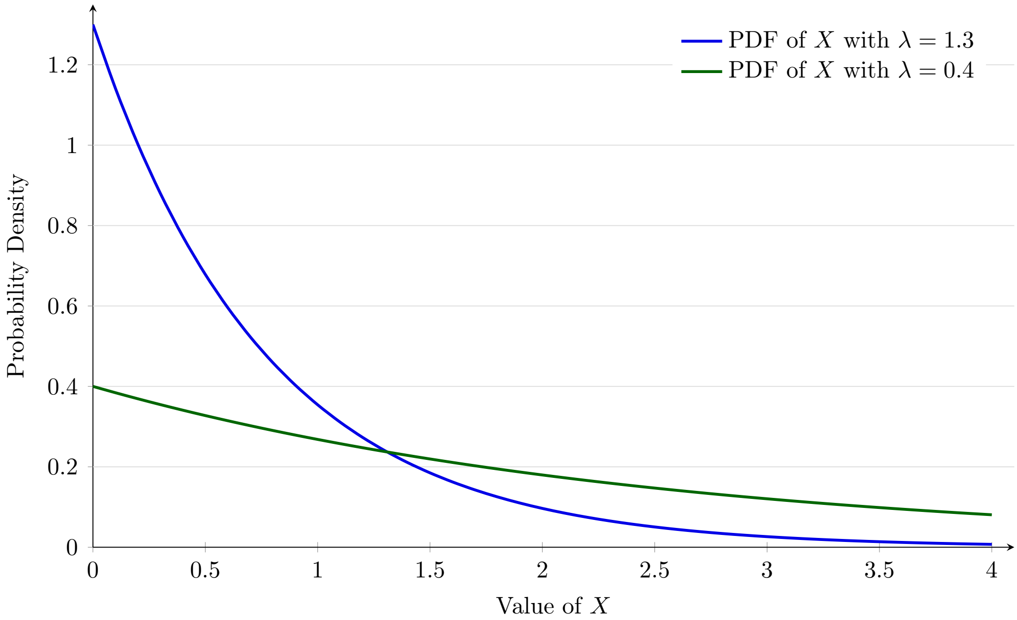


Figure A.1: Probability Density Functions (PDF) of an Exponentially Distributed Random Variable X with Parameter $\lambda = 1.3$ and $\lambda = 0.4$, respectively.

Note: The underlying data was generated in Mathematica.

the given example, it is defined as the conditional probability that a traffic light, having survived t days, will fail. Formally, the failure rate, $r(t)$, is thus given by (Ross, 2010, 299)

$$r(t) = \frac{f(t)}{1 - F(t)}.$$

Inserting from Equations (A.1) and (A.2), one immediately sees that the failure rate is constant in the case of an exponentially distributed random variable[69]

$$r(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

The notion of failure rates will be picked up again after the following exposition on Poisson processes. These are a specific form of a counting process. Generally speaking, a counting process $\{N(t), t \geq 0\}$ is a stochastic process, which counts the number of events, $N(t)$, that have happened up until time t (like, for example, the number of cyclists who have crossed a certain bridge until noon). For a Poisson process, the following definition holds (Ross, 2010, 313): A Poisson process is a counting process with rate $\lambda > 0$, if (i) $N(0) = 0$, (ii) the process has independent increments, and (iii) the number of events in any interval

[69] In fact, as Ross (2010, 299-300) demonstrates, the property of memorylessness exists only for random variables that are exponentially distributed.

of length t has a Poisson distribution with mean λt, i.e. $\forall\, s, t \geq 0$

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad n = 0, 1, \ldots$$

In the context of the model in the main text, for an individual about to engage in research and calculating the value of an innovation, information on the absolute number of innovations in a given sector is of minor interest compared to the time span over which she will be able to earn monopoly profits. Hence, information on the time between innovations is of central interest. Denoting the point of time of the first innovation as T_1 and defining T_n as the time span or interarrival time between event (or innovation) $n-1$ and n, implies that if, for example, innovation number 5 occurred at time 33, innovation 6 at time 34, and the next innovation at time 38, then one would have $T_6 = 34-33 = 1$ and $T_7 = 38 - 34 = 4$ as the values for the interarrival times. Information on the distribution of this sequence of random variables can now be derived by noting that the probability that the first event or innovation occurs after time t is given by the expression (Ross, 2010, 317)

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}.$$

Hence, T_1 is exponentially distributed. This result follows from using property (i) of the Poisson process and by noting that during the interval $[0, t]$ by definition no event occurs so that the number of events in this particular interval is $N(t+0) = N(t) = 0$.[70] Next, the probability of T_2, i.e. the probability that the time between events 1 and 2 is larger than t is the probability of T_2 given that T_1 already happened (which necessarily needs to be the case given the definition of T_2) and had an interarrival time of e.g. s. Then, it holds that

$$\begin{aligned} P\{T_2 > t | T_1 = s\} &= P\{0 \text{ events in } (s, s+t] | T_1 = s\} \\ &= P\{0 \text{ events in } (s, s+t]\} \\ &= e^{-\lambda t}, \end{aligned}$$

where the second line follows from the fact that the Poisson process has independent increments (i.e. the number of events that occur in non-overlapping intervals are independent from each other) so that the conditional and unconditional probabilities are identical. The third line holds, as the process has stationary increments,[71] implying that the distribution for the number of events in $(s, s+t)$ is identical for all s (Ross, 2010,

[70] In general, it holds that $P\{X > x\} = 1 - P\{x \leq X\}$. Hence, with reference to the cumulative distribution function in Equation (A.2), the claim that T_1 has an exponential distribution is valid.

[71] This is implied by an alternative definition of a Poisson process to the one provided above. See Ross (2010, 314) for the details concerning this definition.

313).

Having demonstrated that the interarrival times are independently and identically distributed, the failure rate for the interarrival times is therefore given by the parameter or intensity of the Poisson process, λ. Translated into the context of the model, a failure is equivalent to a new innovation, and the probability that a new innovation comes into existence during the interval dt is given by $r(t)dt = \lambda dt$ (Ross, 2010, 299).

Figures A.2 and A.3 illustrate important characteristics of Poisson processes and the exponential distribution with different values for λ. One clearly sees from the length of the horizontal lines in Figure A.2, that the time interval between innovations (or "failures") is not constant. Interpreting the units of time as years, it takes, for instance, only approximately three months to go from quality level 7 to level 10 (or come up with 3 additional products in that time span), whereas making the three steps from 3 to 6 takes approximately 5 years. Also, the number of absolute innovations is higher for the process with a higher value for λ (17 versus 4).

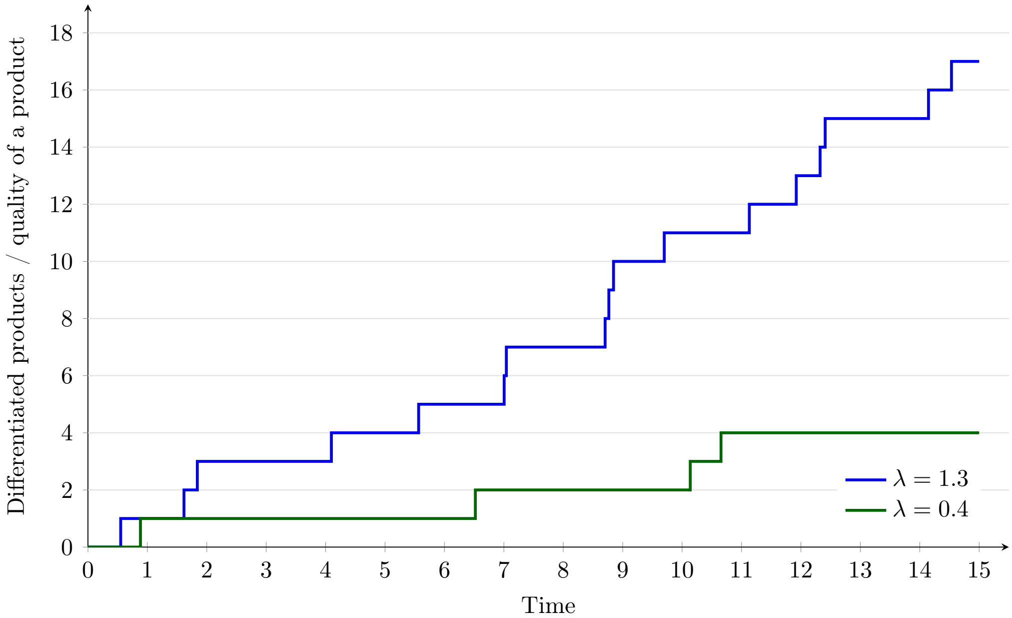


Figure A.2: Illustration of Two Poisson Processes with Intensities $\lambda = 1.3$ and 0.4, respectively.

Note: The data underlying the Poisson processes was generated in Mathematica.

Additionally, Figure A.3 illustrates that a higher value for λ is equivalent to having a larger probability mass at any value of the random variable. Therefore, the probability that an innovation occurs within a certain period of time is indeed higher for higher values of λ.

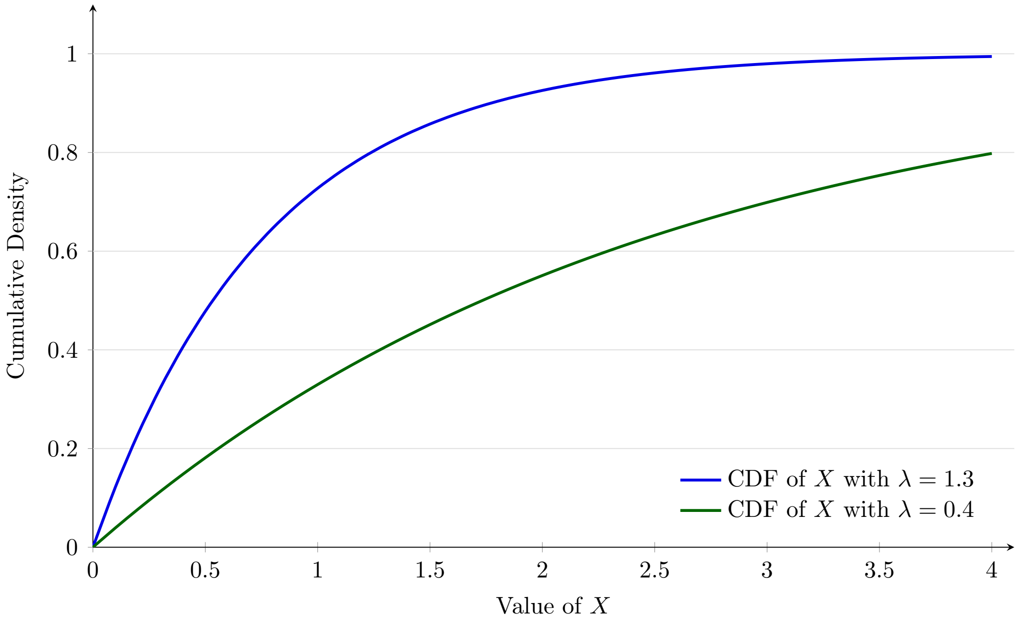


Figure A.3: Cumulative Density Functions (CDF) of an Exponentially Distributed Random Variable X with Parameter $\lambda = 1.3$ and $\lambda = 0.4$, respectively.

Note: The underlying data was generated in Mathematica.

B Additional Derivations – Model

This appendix gathers a variety of derivations of (intermediate) results that are merely stated in the presentation of the model in Sections 2 and 3 in the main text.

B.1 Derivation of the Inverse Demand Schedule for Intermediate Goods

Below, the inverse demand function for an intermediate good will be derived in detail.

The necessary condition for the maximization problem in Equation (2) is given by $\frac{d\Pi_i(v,t)}{dx_i(v,t)} = 0$. Calculating the derivative in this equation, requires an application of the following result for differentiating under the integral sign (see, for instance, Sydsæter et al., 2008, 159)

$$F(x) = \int_c^d f(x,t)\,dt \qquad \Longrightarrow \qquad F'(x) = \int_c^d \frac{\partial f(x,t)}{\partial x}\,dt.$$

Applying this general result to the problem in (2), leads to the necessary condition

$$Q_i^{\alpha-1} \int_0^{Q_i(t)} \alpha A_i(v,t) L_i(t)^{1-\alpha} x_i(v,t)^{-(1-\alpha)}\,dv = \int_0^{Q_i} p_i(t)\,dv.$$

Note that this expression is not identical to the solution given in Equation (3) in the main text. The reason is that in the equation above the derivation has been taken with respect to $x_i(v,t)$ in general (i.e. the whole continuum of varieties) and not with respect to a specific intermediate good, like, for example, good j. For a specific good j, the correct derivative to take is $\frac{\partial \Pi_i(v,t)}{\partial x_i(j,t)}$. This derivative can be stated, by rewriting the integrals in the optimization problem (maybe slightly informally interpreting the integral as a sum of discrete varieties), as

$$
\begin{aligned}
\frac{\partial \Pi_i(v,t)}{\partial x_i(j,t)} = Q_i(t)^{\alpha-1} & \left[\int_{\substack{v \neq j \\ v=0}}^{Q_i(t)} \frac{\partial}{\partial x_i(j,v)} A_i(v,t) x_i(v,t)^{\alpha} L_i(t)^{1-\alpha} dv \right. \\
& \left. + \frac{\partial}{\partial x_i(j,v)} A_i(j,t) x_i(j,t)^{\alpha} L_i(t)^{1-\alpha} \right] \\
& - \left[\int_{\substack{v \neq j \\ v=0}}^{Q_i(t)} p_i(v,t) x_i(v,t) dv + \frac{\partial}{\partial x_i(j,v)} p_i(j,t) x_i(j,t) \right] \\
& - \frac{\partial}{\partial x_i(j,v)} w_i(t) L_i(t).
\end{aligned}
$$

Calculating the respective derivatives in this expression and setting the result equal to zero, leads to (note that the terms with the integrals no longer depend on $x_i(j,t)$ and thus their derivative with respect to this variable is equal to zero)

$$
Q_i(t)^{\alpha-1} \alpha A_i(j,t) L_i(t)^{1-\alpha} x_i(j,t)^{-(1-\alpha)} = p_i(j,t).
$$

As good j is just one specific good out of the continuum $v \in [0, Q_i(t)]$, the (inverse) demand from the producers of final goods for intermediate goods in Equation (3) in the main text follows.

B.2 Deriving the Production Function in Intensive Form

In equilibrium, capital supply, $K_i(t)$, equals capital demand, $\int_0^{Q_i(t)} K_i(v,t) dv$, and hence, substituting $K_i(v,t) = A_i(v,t) x_i(v,t)$ from the production function for intermediate goods (see Equation 5) into this equality, leads to

$$
K_i(t) = \int_0^{Q_i(t)} A_i(v,t) x_i(v,t) \, dv \qquad \Longleftrightarrow \qquad K_i(t) = x_i(t) \int_0^{Q_i(t)} A_i(v,t) \, dv \tag{B.1}
$$

where the second equation has used the property that the equilibrium in the intermediate goods sector is symmetric (see Equation (9)). Defining the average productivity parameter in the intermediate goods sector as

$$A_i(t) \equiv \frac{1}{Q_i(t)} \int_0^{Q_i(t)} A_i(v,t)\, dv \qquad \Longleftrightarrow \qquad A_i(t)Q_i(t) = \int_0^{Q_i(t)} A_i(v,t)\, dv \tag{B.2}$$

and substituting from the second equation in the expression above into Equation (B.1) results in $x_i(t) = \hat{k}_i(t)\frac{L_i(t)}{Q_i(t)}$, where $\hat{k}_i(t)$ is the capital stock per effective worker, i.e. $\hat{k}_i(t) \equiv \frac{K_i(t)}{A_i(t)L_i(t)}$. With the property of the symmetric equilibrium, the expression for $x_i(t)$ just derived, and the second expression in Equation (B.2), the production function in intensive form can be written as

$$\hat{y}_i(t) = \hat{k}_i(t)^\alpha$$

where $\hat{y}_i(t) \equiv \frac{Y_i(t)}{A_i(t)L_i(t)}$ is the production per effective worker, and which is identical to Equation (4).

In the main text in Section 2.1, it was stated that as the production function is multiplied by the factor $Q_i(t)^{\alpha-1}$, technological progress in this model is due to increases in productivity and not increases in the number of varieties as in the model by Romer (1990). This result will now be demonstrated mathematically. The production function in intensive form above can be expressed in aggregate terms as $Y_i(t) = A_i(t)L_i(t)^{1-\alpha}K_i(t)^\alpha A_i(t)^{-\alpha}$. From the two expressions after the equivalence arrows in Equations (B.1) and (B.2), it follows that $\frac{K_i(t)}{A_i(t)} = x_i(t)Q_i(t)$ so that

$$Y_i(t) = A_i(t)L_i(t)^{1-\alpha}\big(x_i(t)Q_i(t)\big)^\alpha \tag{B.3}$$

which has constant returns to scale in the two input factors labor and aggregate amount of intermediate inputs. As can be seen, technological progress in this specification is only due to increases in productivity. On the other hand, without the factor $Q_i(t)^{\alpha-1}$ in the original specification of the production function in Equation (1), the right-hand side in Equation (B.3) would need to be multiplied by $Q_i(t)^{1-\alpha}$ to obtain a corresponding result, and increases in the number of varieties would lead to increases in productivity in this case.

B.3 Convergence of the Number of Workers per Product to a Constant

The result that the number of workers per intermediate good, $L_i(t)/Q_i(t) = l_i(t)$ monotonically converges to the constant n_i/ξ_i in Equation (13) can be derived as follows:

Taking the natural logarithm of $l_i(t)$ and deriving the result with respect to time yields

$$\frac{\dot{l}_i(t)}{l_i(t)} = \frac{\dot{L}_i(t)}{L_i(t)} - \frac{\dot{Q}_i(t)}{Q_i(t)}.$$

Inserting n_i for the population growth rate and substituting for $\dot{Q}_i(t)$ from Equation (12) leads to the differential equation

$$\frac{\dot{l}_i(t)}{l_i(t)} = n_i - \xi l_i(t) \qquad \Longleftrightarrow \qquad \dot{l}_i(t) - n_i l_i(t) = -\xi l_i(t)^2. \tag{B.4}$$

This equation has one steady state at $l_i^* = \frac{n_i}{\xi}$, which results from setting $\dot{l}_i(t) = 0$ and solving for l_i. Asymptotic convergence to the steady state follows as $\dot{l}_i(t) < 0$ for all $l_i(t) > l_i^*$ and $\dot{l}_i(t) > 0$ for all $l_i(t) < l_i^*$ (see Part 3 of Corollary 2.2 in Acemoglu (2009) for this approach to determine global asymptotic stability).[72]

B.4 Derivation of the Value of an Innovation

This section provides a derivation of the expression for the value of an innovation to a monopolist stated in Equation (16). In particular, it will be shown how this value depends on the Poisson arrival rate of new (quality) innovations.

A firm will reap monopoly profits from the time the innovation is brought to market (e.g. $t = 0$) until it is replaced at some time T, with $T \in (0, \infty)$, by a new monopolist producing a variety of a higher quality, and profits will fall to zero.[73] Therefore, the value for a firm at time 0 is given by[74]

$$V_d(0) = \int_0^T e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau$$

[72]Note that Equation (B.4) is a Bernoulli equation (Sydsæter et al., 2008, 208), which can be transformed into a standard linear differential equation by using the transformation $z(t) = \frac{1}{l_i(t)}$ and then be solved for the general solution

$$z(t) = \left(z(0) - \frac{\xi}{n_i}\right) e^{-n_i t} + \frac{\xi}{n_i}.$$

Reversing the transformation, the general solution for $l_i(t)$ is thus given by

$$l_i(t) = \frac{1}{\left(l_i(0)^{-1} - \frac{\xi}{n_i}\right) e^{-n_i t} + \frac{\xi}{n_i}},$$

which confirms that $l_i^* = \frac{n_i}{\xi}$ is indeed a steady-state value for the differential equation.

[73]This follows from the Arrow replacement effect and the fact that the previous monopolist will be driven out of the market via Bertrand competition, as the new innovator produces a higher quality good at identical costs (Aghion et al., 2014, 518).

[74]The subscript d denotes "deterministic" in this instance.

where $r(s)$ is the interest rate at time s, and the exponential expression is the discount factor applied to the monopolist's profits. That the replacement will happen is certain, but the point in time T in the future when it will happen can only be determined with some probability. Hence, the expected value of an innovation is a random variable and can be expressed as follows[75]

$$\begin{aligned} V(0) = E[V_d(0)] &= \int_0^\infty f(T) \left[\int_0^T e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau \right] dT \\ &= \int_0^\infty \int_0^T f(T) e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau\, dT, \end{aligned} \tag{B.5}$$

where $f(T)$ is a general probability density function with $f(T) \geq 0 \,\forall\, T$ and $\int_0^\infty f(T)\, dT = 1$. The equality in Equation (B.5) follows as $f(T)$ does not depend on τ and can thus be moved into the integral with respect to τ. However, this expression is still quite different from Equation (16).

The next step is to change the order of integration, which requires adjusting the limits of integration (this step is explained and illustrated in more detail at the end of this section). This procedure yields

$$V(0) = \int_0^\infty \left[\int_\tau^\infty f(T)\, dT \right] e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau. \tag{B.6}$$

Referring back to the discussion on Poisson processes in Appendix A, and making a specific distributional assumption on the function $f(T)$ (compare Equation (A.1)), the integral in brackets is just the probability that an innovation occurs after time τ, or, equivalently, that the firm can still earn monopoly profits at time τ. Calculating the complementary probability to the one stated (in general terms) in Equation (A.2), this probability is $e^{-\phi\tau}$ so that the value of an innovation is given by[76]

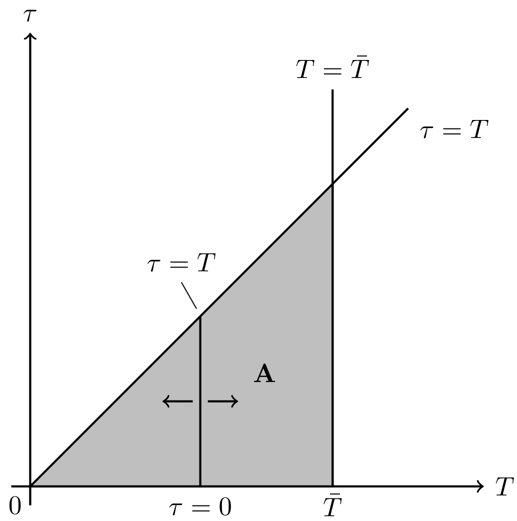
$$\begin{aligned} V(0) &= \int_0^\infty e^{-\phi\tau} \cdot e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau \\ &= \int_0^\infty e^{-\int_0^\tau (r(s)+\phi)ds} \pi(\tau)\, d\tau. \end{aligned}$$

To be precise, this expression differs slightly from the more general one in the main text, as it is assumed here that ϕ is constant, which only holds in steady state (also, an identifier i for individual regions was dropped here).

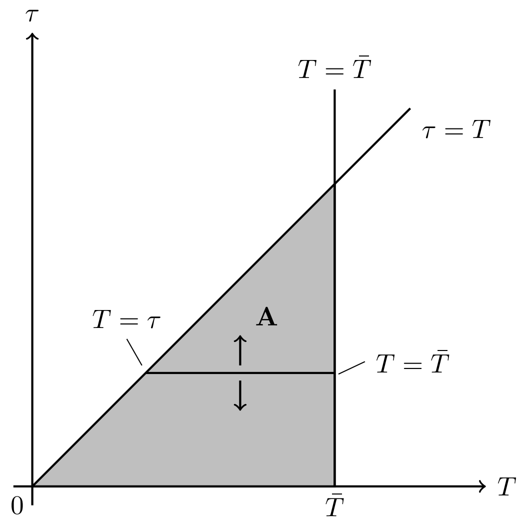
[75]Note that this approach is basically the same as the one adopted by Yaari (1965, 142) in his model of uncertain lifetime.

[76]The second equality follows, as $e^{-\int_0^\tau \phi\, ds} = e^{-\left.\phi s\right|_0^\tau} = e^{-\phi\tau}$.

Proving the validity of the change in the order of integration above, requires demonstrating that the expressions in Equation (B.5) and (B.6) are equivalent. This basically works by showing that the area of integration is identical in both cases. The following method can, for instance, be found in Sydsæter et al. (2008, 166pp) or Thomas Jr. (2005, 1074-75). In general, double integrals are evaluated by first working out the inner integral and then the outer one. In illustrating the method graphically in Figure B.4, the upper limit of integration in Equation (B.5) is changed from ∞ to an upper bound of $\bar{T}$ to simplify the graphical exposition. For the double integral in Equation (B.5), the relevant area of integration is depicted in Panel (a) of Figure B.4 and the one for Equation (B.6) in Panel (b).



(a) Illustration of Evaluating the Integral in Equation (B.5).



(b) Illustration of Evaluating the Integral in Equation (B.6).

Figure B.4: Graphical Illustration of Changing the Order of Integration and Preserving the Area of Integration.

In order to evaluate the integral in the first equation, the inner integral is evaluated along the line $\tau = 0$ to $\tau = T$, and then the outer integral is evaluated by integrating along all vertical lines from $T = 0$ to $T = \bar{T}$ (indicated by the horizontal arrows) to obtain the grey-shaded area of integration **A** (see Panel (a)). The same area is obtained by changing the order of integration, then first integrating along the horizontal line from $T = \tau$ to $T = \bar{T}$, and then the outer integral covers all horizontal lines from $\tau = 0$ to $\tau = \bar{T}$ (indicated by the vertical arrows) so that also in this case the grey-shaded area of integration **A** is obtained (see Panel (b)).

B.5 Derivation of the Research-Arbitrage Equation

The research-arbitrage equation results from deriving Equation (17) which is repeated here for convenience

$$v_i(t) = \int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\,d\tau$$

with respect to t. Taking this derivative requires applying Leibniz's Formula (see, for example Sydsæter et al., 2008, 160)

$$\begin{aligned} F(t) &= \int_{u(t)}^{v(t)} f(t,\tau)\,d\tau \\ \implies \quad F'(t) &= \underbrace{f\big(t,v(t)\big)}_{①} \underbrace{v'(t)}_{②} - \underbrace{f\big(t,u(t)\big)}_{③} \underbrace{u'(t)}_{④} + \int_{u(t)}^{v(t)} \underbrace{\frac{\partial f(t,\tau)}{\partial t}}_{⑤}\,d\tau. \end{aligned}$$

In the case at hand, the function $f(t,\tau)$ in the formula above is therefore given by $f(t,\tau) = e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)$, and the individually numbered terms above are given by the following expressions, respectively

①: $f\big(t,v(t)\big) = e^{-\int_t^\infty (r_i(s)+\phi_i(s))\,ds} \lim_{\tau\to\infty} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)$

②: $v'(t) = \frac{d}{dt}\text{“}\infty\text{”} = 0$

③: $f\big(t,u(t)\big) = e^{-\int_t^t (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(t)\big) l_i(t) = e^0 \tilde{\pi}_i\big(\hat{k}_i(t)\big) l_i(t) = \tilde{\pi}_i\big(\hat{k}_i(t)\big) l_i(t)$

④: $u'(t) = \frac{d}{dt} t = 1$

⑤: $\frac{\partial}{\partial t} e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau) = -[-r_i(t)+\phi_i(t)] e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)$

Using these intermediate results with the product ① · ② $= 0$ already inserted, it follows that

$$\begin{aligned} \frac{\partial}{\partial t} v_i(t) = \dot{v}_i(t) &= -\tilde{\pi}_i\big(\hat{k}_i(t)\big) l_i(t) + \int_t^\infty [r_i(t)+\phi_i(t)] e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\,d\tau \\ &= -\tilde{\pi}_i\big(\hat{k}_i(t)\big) l_i(t) + [r_i(t)+\phi_i(t)] \underbrace{\int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\,d\tau}_{v_i(t)} \\ &= -\tilde{\pi}_i\big(\hat{k}_i(t)\big) l_i(t) + [r_i(t)+\phi_i(t)] v_i(t) \end{aligned}$$

and from the last expression, the research-arbitrage stated in the main text in Equation (17), is readily obtained.

B.6 Convergence of Relative Productivities

In the following, it will be demonstrated that the relative productivity parameters $a_i(v,t) = \frac{A_i(v,t)}{A_i(t)^{max}}$ converge to an invariant distribution. More specifically, it will be shown that the distribution of the fraction of sectors for which $A_i(v,t) \leq A_i(t)^{max}$ is time independent and given by $a_i^{\frac{1}{\sigma}}$. This result is based on the assumption that new and existing products have identical distributions for the productivity parameters at any time t. The proof follows along the lines of Aghion and Howitt (1998, 115).

For an arbitrary point in time t, denote the cumulative distribution of the absolute productivity parameters by $F(\cdot,t)$. At some point in time, $t_0 \geq 0$, one particular sector $v \in [0, Q_i(t)]$ with productivity parameter $A_i(v,t)$ necessarily was the leading-edge sector. Defining then the cumulative distribution function as $\Phi_i(t) = F\big(A_i(v,t),t\big)$, it needs to hold that

$$\Phi_i(t_0) = 1, \tag{B.7}$$

i.e. the probability that the particular sector that was picked out has the highest productivity across all sectors under consideration equals 1. At time t_0 "many" sectors are behind the one with the highest productivity. These sectors individually will innovate with the Poisson arrival rate for vertical innovations and, hence, in aggregate, since there are $\Phi_i(t)$ sectors, with the rate $\Phi_i(t)\lambda_i\kappa_i(t)^\phi$. This rate therefore equals the one with which the mass of sectors behind the leading one will decrease. In formal terms,

$$\dot{\Phi}_i(t) = -\Phi_i(t)\lambda_i\kappa_i(t)^\phi \qquad \forall\, t \geq t_0. \tag{B.8}$$

Equations (B.7) and (B.8) pose then an initial-value problem with solution

$$\Phi_i(t) = e^{-\int_{t_0}^{t} \lambda_i\kappa_i(s)^\phi\, ds} \qquad \forall\, t \geq 0. \tag{B.9}$$

Equation (20) implies the differential equation $\dot{A}_i(t)^{max} = \sigma\lambda_i\kappa_i(t)^\phi A_i(t)^{max}$, and at the start of this section it was assumed that $A_i(v,t) = A_i(t_0)^{max}$ (compare also the definition in Equation (15)). The solution to the differential equation for the leading-edge productivity parameter is therefore

$$A_i(t)^{max} = A_i(v,t)e^{\sigma\int_{t_0}^{t} \lambda_i\kappa_i(s) ds} \qquad \forall\, t \geq t_0. \tag{B.10}$$

From combining Equations (B.9) and (B.10), it thus follows that the distribution of the relative productivities in the long run converges to

$$\Phi_i(t) = \left(\frac{A_i(v,t)}{A_i(t)^{max}}\right)^{\frac{1}{\sigma}} = a_i(t)^{\frac{1}{\sigma}}.$$

As Aghion and Howitt (1992, 116) point out, in the long run almost all values for a_i in the interval $[0,1]$ will exist.[77]

B.7 Growth Rate of the Average and Leading-Edge Productivity Parameters

In the main text at the end of Section 2.3, an equation for the growth rate of the average productivity, $A_i(t)$, was given, which is repeated here for convenience

$$\dot{A}_i(t) = \lambda_i \kappa_i(t)^{\phi}\big(A_i(t)^{max} - A_i(t)\big).$$

It will now be demonstrated that the leading-edge and average productivity parameters will grow at identical rates. Defining the ratio between these two parameters as $\Gamma_i \equiv \frac{A_i(t)^{max}}{A_i(t)}$ and rewriting it in growth rates leads to

$$\frac{\dot{\Gamma}_i(t)}{\Gamma_i(t)} = \frac{\dot{A}_i(t)^{max}}{A_i(t)^{max}} - \frac{\dot{A}_i(t)}{A_i(t)}. \tag{B.11}$$

Substituting for the growth rate of the leading-edge parameter from Equation (20) and noting that the growth rate of the average productivity parameter is given by

$$\frac{\dot{A}_i(t)}{A_i(t)} = \lambda_i \kappa_i(t)^{\phi}\left[\frac{A_i(t)^{max}}{A_i(t)} - \frac{A_i(t)}{A_i(t)}\right] = \lambda_i \kappa_i(t)^{\phi}\big(\Gamma_i(t) - 1\big),$$

it follows that the growth rate of the ratio of the productivity parameters is

$$\frac{\dot{\Gamma}_i(t)}{\Gamma_i(t)} = \sigma\lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi}\left[\Gamma_i(t) - 1\right].$$

This expression can be rewritten as[78]

$$\dot{\Gamma}_i(t) = \big[(1+\sigma)\lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi}\Gamma_i(t)\big]\,\Gamma_i(t) \tag{B.12}$$

[77] Additional remarks on the cross-section distribution, including a graphical analysis can be found in Howitt (2000, 834). See also Howitt (1999, 721).

[78] As in the case for the differential equation for the number of workers, $l_i(t)$, Equation (B.12) is also a Bernoulli equation.

which has a trivial steady state at zero and a second one at $\Gamma_i^* = 1+\sigma$. As long as $\lambda_i \kappa_i(t)^{\phi} > 0$, convergence to this value follows via applying the same approach as in Appendix B.3. From the definition of Γ_i, it holds that $A_i(t)^{max} = (1+\sigma)A_i(t)$, and both productivity parameters will therefore grow at the rate $g_i(t) = \sigma\lambda_i\kappa(t)^{\phi}$.

B.8 Derivation of the Global Technology Growth Rate

This section derives the productivity growth rate given in Equation (26) in the main text. Starting with inserting the expression for κ_i in Equation (25) into the one for g^w in Equation (24), yields

$$g^w = \frac{\sigma\lambda}{[(1+\sigma)\xi]^{\phi}} s_{A,i}^{\phi} y_i^{\phi} n_i^{\phi} A_i^{-\phi} \prod_{j=1}^{N} \left(\frac{A_j}{A_i}\right)^{\gamma_i v_{ij}}. \tag{B.13}$$

With the help of the properties of the product operator, the last factor can now be rewritten in the following way

$$\begin{aligned}
\prod_{j=1}^{N} \left(\frac{A_j}{A_i}\right)^{\gamma_i v_{ij}} &= \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} A_i^{-\gamma_i v_{ij}} = \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} \prod_{j\neq i}^{N} A_i^{-\gamma_i v_{ij}} \\
&= \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} A_i^{-\gamma_i \sum_{j\neq i}^{N} v_{ij}} \\
&= \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} A_i^{-\gamma_i (1-v_{ii})} \\
&= A_i^{-\gamma_i \frac{1}{\gamma_i}} \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} = A_i^{-1} \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}}.
\end{aligned}$$

The step from the first to the second line uses the result $\sum_{j=1}^{N} v_{ij} = 1$ and the one from the second-to-last to the last line takes advantage of the definition $v_{ii} \equiv \frac{\gamma_i - 1}{\gamma_i} < 1$ (see Ertur and Koch (2011, 226-27) on these assumptions). Substituting the final result in the derivation above into Equation (B.13), leads to Equation (26) in Section 3.1.

B.9 Existence of $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$

In contrast to the case of a row-standardized interaction matrix, $\boldsymbol{I} - \gamma\boldsymbol{W}$ might be singular also for values in the interval $\gamma \in (-1, 1)$. The general condition for this matrix to be singular is $|\boldsymbol{I} - \gamma\boldsymbol{W}| = 0$, i.e. if $\frac{1}{\gamma}$ is an eigenvalue of the interaction matrix. Consider now, for instance, the matrix

$$\boldsymbol{W}_1 = \begin{pmatrix} 0 & 16 \\ 4 & 0 \end{pmatrix},$$

which is not row-standardized. Its characteristic equation is given by $\lambda^2 = 64$ so that the eigenvalues are $\lambda_1 = -8$ and $\lambda_2 = 8$. Then, for $\gamma = \frac{1}{8}$ the matrix $\boldsymbol{I} - \gamma \boldsymbol{W}_1$ will be singular. However, by restricting the parameter space for γ to $\gamma \in \left(-\frac{1}{\lambda_1}, -\frac{1}{\lambda_2}\right)$, the inverse above will be non-singular. An equivalent representation of the model under consideration can thus be obtained if the interaction matrix is normalized by this factor, i.e. $\boldsymbol{W}_1^* = \frac{\boldsymbol{W}_1}{\lambda_2}$ and by denoting $\gamma^* = \gamma\lambda_2$ with parameter space $\gamma^* \in (-1, 1)$. A similar procedure works in more general cases (Kelejian and Prucha, 2010, 56), when the eigenvalues cannot be as easily determined as in the matrix above. With the help of Gerschgorin's Circle Theorem (Gerschgorin, 1931), regions in the complex plane can be determined that contain the eigenvalues of the matrix.[79] With this information, it is possible to identify an interval for the parameter space, in which the inverse exists (see also Ertur and Koch, 2011, 231).

[79]A more recent formal statement of this theorem can, for example, be found in Cheney and Kincaid (2008).

C Additional Derivations – Econometric Theory

This appendix derives results in detail that are important in the econometric estimation of the model and for drawing inference. More specifically, a part of the score vector will be derived, before the steps in the derivation of the variance-covariance matrix, which is merely stated in Ertur and Koch (2011, 233), will be demonstrated.

C.1 Derivation of the Maximum Likelihood Estimator $\hat{\boldsymbol{\delta}}$

Before taking the derivative of the likelihood function with respect to $\boldsymbol{\delta}$, it will first be written in an expanded form. From Equation (43), it follows that

$$\begin{aligned}
\ln L(\boldsymbol{y};\boldsymbol{\delta},\gamma,\sigma^2) &= \underbrace{-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|\boldsymbol{I}-\gamma\boldsymbol{W}|}_{\equiv C} \\
&\quad - \frac{1}{2\sigma^2}\left[\boldsymbol{y}'(\boldsymbol{I}-\gamma\boldsymbol{W})'(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y} - \underbrace{\boldsymbol{y}'(\boldsymbol{I}-\gamma\boldsymbol{W})'\tilde{\boldsymbol{X}}\boldsymbol{\delta}}_{1\times 1}\right. \\
&\qquad \left. - \underbrace{\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}}_{1\times 1} + \boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right] \\
&= C - \frac{1}{2\sigma^2}\left[\boldsymbol{y}'(\boldsymbol{I}-\gamma\boldsymbol{W})'(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y} - 2\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y} + \boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]
\end{aligned}$$

where the last line has used the fact that the terms with the underbraces are identical scalars and that the matrix $(\boldsymbol{I}-\gamma\boldsymbol{W})$ is symmetric. Taking now the derivative[80] with respect to $\boldsymbol{\delta}$ leads to

$$\frac{\partial \ln L(\cdot)}{\partial \boldsymbol{\delta}} = -\frac{1}{2\sigma^2}\left[-2\tilde{\boldsymbol{X}}'(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y} + 2\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right].$$

Setting this expression equal to zero and solving for $\hat{\boldsymbol{\delta}}$, yields the expression in Equation (47).

C.2 Derivation of the Variance-Covariance Matrix

The asymptotic variance-covariance matrix is given by the inverse of the information matrix $\boldsymbol{I}(\boldsymbol{\delta},\gamma,\sigma^2)$, and this matrix is equal to the negative expected Hessian matrix, $\boldsymbol{H}$,

[80] For the rules on matrix derivation see, for example, Verbeek (2004, 394-95).

for the log-likelihood function in Equation (43). In general terms, the information matrix thus reads

$$\boldsymbol{I}(\boldsymbol{\delta},\gamma,\sigma^2) = -E[\boldsymbol{H}] = -E\begin{bmatrix} \frac{\partial^2 \ln L(\cdot)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} & \frac{\partial^2 \ln L(\cdot)}{\partial\boldsymbol{\delta}\partial\gamma} & \frac{\partial^2 \ln L(\cdot)}{\partial\boldsymbol{\delta}\partial\sigma^2} \\ \frac{\partial^2 \ln L(\cdot)}{\partial\gamma\partial\boldsymbol{\delta}'} & \frac{\partial^2 \ln L(\cdot)}{\partial\gamma^2} & \frac{\partial^2 \ln L(\cdot)}{\partial\gamma\partial\sigma^2} \\ \frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\boldsymbol{\delta}'} & \frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\gamma} & \frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\sigma^2} \end{bmatrix}. \tag{C.1}$$

The individual entries for the first row in the Hessian matrix are calculated by taking the respective partial derivatives of Equation (C.1):

$$\frac{\partial^2 \ln L(\cdot)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} = -\frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} \tag{C.2}$$

$$\frac{\partial^2 \ln L(\cdot)}{\partial\boldsymbol{\delta}\partial\gamma} = -\frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y} \tag{C.3}$$

$$\frac{\partial^2 \ln L(\cdot)}{\partial\boldsymbol{\delta}\partial\sigma^2} = \frac{1}{\sigma^4}\left\{-\tilde{\boldsymbol{X}}'\left[\left(I-\gamma\boldsymbol{W}\right)\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]\right\} = -\frac{1}{\sigma^4}\tilde{\boldsymbol{X}}'\boldsymbol{\varepsilon} \tag{C.4}$$

where the last equality has used the expression for $\boldsymbol{\varepsilon}$ in Equation (44).

In order to calculate the entries in the second row of the Hessian, the first derivative of the log-likelihood function with respect to γ is needed. Note that the last term in the log-likelihood function can be equivalently written as $-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$, and the derivative of this term with respect to γ is given by (compare, for example, Anselin, 1988b, 75)

$$\frac{\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial\gamma} = \boldsymbol{\varepsilon}'\frac{\partial\boldsymbol{\varepsilon}}{\partial\gamma} + \frac{\partial\boldsymbol{\varepsilon}'}{\partial\gamma}\boldsymbol{\varepsilon} = 2\boldsymbol{\varepsilon}'\frac{\partial\boldsymbol{\varepsilon}}{\partial\gamma}. \tag{C.5}$$

Deriving the log determinant with respect to γ makes use of Jacobi's formula (compare, for instance, Absil et al., 2008, 196). This states that the derivative of the determinant of a matrix $\boldsymbol{X}$ with respect to a can be expressed in the following way: $\frac{\partial|\boldsymbol{X}|}{\partial a} = \text{tr}\left[\text{adj}\left(\boldsymbol{X}\right)\frac{\partial\boldsymbol{X}}{\partial a}\right]$.[81] Alternatively, provided that $\boldsymbol{X}$ is invertible, the expression for the adjugate matrix, $\text{adj}(\boldsymbol{X}) = |\boldsymbol{X}|(\boldsymbol{X})^{-1}$, can be inserted, implying that the derivative of the determinant is given by $|\boldsymbol{X}|\text{tr}\left[(\boldsymbol{X})^{-1}\frac{\partial\boldsymbol{X}}{\partial a}\right]$. In the following, a derivative of a log determinant will be taken so that taking into account the rules for differentiating logarithmic functions and the ones for determinants, Jacobi's formula reads $\frac{\partial\ln|\boldsymbol{X}|}{\partial a} = \text{tr}\left[(\boldsymbol{X})^{-1}\frac{\partial\boldsymbol{X}}{\partial a}\right]$ in this case.

Applying these rules to the case at hand and noting that $\frac{\partial\boldsymbol{\varepsilon}}{\partial\gamma} = -\boldsymbol{W}\boldsymbol{y}$, the partial derivative of the log-likelihood function with respect to γ is

$$\frac{\partial\ln L(\cdot)}{\partial\gamma} = -\text{tr}\left(\boldsymbol{I}-\gamma\boldsymbol{W}\right)^{-1}\boldsymbol{W} + \frac{1}{\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}.$$

[81] A proof of this result can be found, for instance, in Magnus and Neudecker (1999, 150).

Before calculating the first entry in the second row of the Hessian matrix, note that the expressions $\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}$ and $(\boldsymbol{W}\boldsymbol{y})'\boldsymbol{\varepsilon}$ denote an identical scalar. As $\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}$ is a scalar, it is possible to introduce the trace operator (see Anselin, 1988b, 77) so that

$$\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y} = \mathrm{tr}\left[\boldsymbol{\varepsilon}'(\boldsymbol{W}\boldsymbol{y})\right] = \mathrm{tr}\left\{\left[\boldsymbol{\varepsilon}'(\boldsymbol{W}\boldsymbol{y})\right]'\right\} = \mathrm{tr}\left[(\boldsymbol{W}\boldsymbol{y})'\boldsymbol{\varepsilon}\right] = (\boldsymbol{W}\boldsymbol{y})'\boldsymbol{\varepsilon}'. \tag{C.6}$$

The second equality holds, since a matrix and its transpose have the same trace, and the third equality follows from the properties of transposed matrices. Substituting this expression into the first derivative above, inserting for $\boldsymbol{\varepsilon}$, and taking the partial derivative with respect to $\boldsymbol{\delta}'$ yields the result

$$\frac{\partial^2 \ln L(\cdot)}{\partial\gamma\partial\boldsymbol{\delta}'} = -\frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y}\right)' \tag{C.7}$$

which is just the transpose of Equation (C.3). Anselin (1988b, 75) provides a helpful rule for taking the derivative of an inverse matrix, i.e. $\frac{\partial(\boldsymbol{X})^{-1}}{\partial a} = -(\boldsymbol{X})^{-1}\frac{\partial\boldsymbol{X}}{\partial a}(\boldsymbol{X})^{-1}$ and notes that the trace operator can be applied after differentiation as it is a linear operator. Hence,

$$\begin{aligned}
\frac{\partial^2 \ln L(\cdot)}{\partial\gamma^2} &= -\,\mathrm{tr}\left[-(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}(-\boldsymbol{W})(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\boldsymbol{W}\right] - \frac{1}{\sigma^2}(\boldsymbol{W}\boldsymbol{y})'\boldsymbol{W}\boldsymbol{y} \\
&= -\,\mathrm{tr}\left[\boldsymbol{W}(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\boldsymbol{W}(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\right] - \frac{1}{\sigma^2}(\boldsymbol{W}\boldsymbol{y})'\boldsymbol{W}\boldsymbol{y} \\
&= -\,\mathrm{tr}\left(\boldsymbol{W_A}\boldsymbol{W_A}\right) - \frac{1}{\sigma^2}(\boldsymbol{W}\boldsymbol{y})'\boldsymbol{W}\boldsymbol{y}
\end{aligned} \tag{C.8}$$

where the second equality has taken advantage of the property that the trace of a matrix is invariant to cyclical permutations (see, e.g. Meyer, 2000, 110). Additionally, in the expression in the last equality, the following definition from Ertur and Koch (2011, 233) is employed $\boldsymbol{W_A} \equiv \boldsymbol{W}(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}$. Turning to the last entry in the second row of the Hessian, this is given by

$$\frac{\partial^2 \ln L(\cdot)}{\partial\gamma\partial\sigma^2} = -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}. \tag{C.9}$$

The derivatives in the third row of the Hessian matrix are the partial derivatives of the expression in Equation (45). The first entry in this row is

$$\begin{aligned}
\frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\boldsymbol{\delta}'} &= \frac{1}{2\sigma^4}\left[-2\boldsymbol{y}'(\boldsymbol{I}-\gamma\boldsymbol{W})'\tilde{\boldsymbol{X}} + 2\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right] \\
&= -\frac{1}{\sigma^4}\left[\boldsymbol{y}'(\boldsymbol{I}-\gamma\boldsymbol{W})' - \boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\right]\tilde{\boldsymbol{X}} = -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\tilde{\boldsymbol{X}} = -\frac{1}{\sigma^4}\left(\tilde{\boldsymbol{X}}'\boldsymbol{\varepsilon}\right)'
\end{aligned} \tag{C.10}$$

where again Equation (44) has been used. Applying one more time the rule in Equation

(C.5) for differentiating the expression for the sum of squared errors, facilitates calculating the partial derivative with respect to γ so that

$$\frac{\partial^2 \ln L(\cdot)}{\partial \sigma^2 \partial \gamma} = \frac{1}{2\sigma^4} 2\boldsymbol{\varepsilon'} \left(-\boldsymbol{Wy}\right) = -\frac{1}{\sigma^4}\boldsymbol{\varepsilon' Wy}. \tag{C.11}$$

For the last entry in the third row, the partial derivative reads

$$\frac{\partial^2 \ln L(\cdot)}{\partial \sigma^2 \partial \sigma^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{\varepsilon' \varepsilon}. \tag{C.12}$$

Gathering the results in Equations (C.2) – (C.4) and Equations (C.7) – (C.12) yields the following Hessian matrix of dimension 7×7 (the first column has dimension 7×5 and the remaining two columns each have dimension 7×1)

$$\boldsymbol{H} = \begin{pmatrix} -\frac{1}{\sigma^2}\boldsymbol{\tilde{X}'\tilde{X}} & -\frac{1}{\sigma^2}\boldsymbol{\tilde{X}'Wy} & -\frac{1}{\sigma^4}\boldsymbol{\tilde{X}'\varepsilon} \\ -\frac{1}{\sigma^2}\left(\boldsymbol{\tilde{X}'Wy}\right)' & -\mathrm{tr}\left(\boldsymbol{W_A W_A}\right) - \frac{1}{\sigma^2}\left(\boldsymbol{Wy}\right)'\boldsymbol{Wy} & -\frac{1}{\sigma^4}\boldsymbol{\varepsilon' Wy} \\ -\frac{1}{\sigma^4}\left(\boldsymbol{\tilde{X}'\varepsilon}\right)' & -\frac{1}{\sigma^4}\boldsymbol{\varepsilon' Wy} & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{\varepsilon'\varepsilon} \end{pmatrix}.$$

The next step in deriving the information matrix is taking the (negative) expected value of the Hessian matrix above. Starting with the first column, its first entry contains no random variables and hence $-E\left[-\frac{1}{\sigma^2}\boldsymbol{\tilde{X}'\tilde{X}}\right] = \frac{1}{\sigma^2}\boldsymbol{\tilde{X}'\tilde{X}}$. Moving on, the second entry equals

$$\begin{aligned} -E\left[-\frac{1}{\sigma^2}\left(\boldsymbol{\tilde{X}'Wy}\right)'\right] &= \frac{1}{\sigma^2}\left\{E\left[\boldsymbol{\tilde{X}'W}(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\boldsymbol{\tilde{X}\delta} + (\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}\right]\right\}' \\ &= \frac{1}{\sigma^2}\left(\boldsymbol{\tilde{X}'W_A\tilde{X}\delta}\right)' \end{aligned} \tag{C.13}$$

where the last line follows as the expectation is a linear operator, the errors are assumed to be independent of all explanatory variables, and since $E[\boldsymbol{\varepsilon}] = 0$ due to the distributional assumption from Section 4.1. These latter two results can also be applied to calculate the last entry in the first column, implying that $-E\left[-\frac{1}{\sigma^4}\left(\boldsymbol{\tilde{X}'\varepsilon}\right)'\right] = \boldsymbol{0}$.

In the second column of the information matrix, the first entry is simply the transpose of the 1×5 vector in Equation (C.13). However, the second entry on the diagonal requires more computations. The negative of the expected value of the first term in this entry is not a random variable and thus equals $-E\left[-\mathrm{tr}\left(\boldsymbol{W_A W_A}\right)\right] = \mathrm{tr}\left(\boldsymbol{W_A W_A}\right)$, while the following holds for the second term

$$
\begin{aligned}
-E\left[-\frac{1}{\sigma^2}\left(\boldsymbol{W}\boldsymbol{y}\right)'\boldsymbol{W}\boldsymbol{y}\right] &= \frac{1}{\sigma^2}E\left\{\left[\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}+\boldsymbol{W_A}\boldsymbol{\varepsilon}\right]'\left[\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}+\boldsymbol{W_A}\boldsymbol{\varepsilon}\right]\right\} \\
&= \frac{1}{\sigma^2}\left\{E\left[\left(\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]+E\left[\left(\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)'\boldsymbol{W_A}\boldsymbol{\varepsilon}\right]\right. \\
&\quad \left. +E\left[\left(\boldsymbol{W_A}\boldsymbol{\varepsilon}\right)'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]+E\left[\left(\boldsymbol{W_A}\boldsymbol{\varepsilon}\right)'\boldsymbol{W_A}\boldsymbol{\varepsilon}\right]\right\}.
\end{aligned}
$$

Following the same arguments as above, the first term in the previous equation is completely deterministic and the cross products have an expected value of $\mathbf{0}$. For the last term, the rules from Equation (C.6) and the fact that cyclical permutations leave the trace of a matrix unchanged can be applied to demonstrate that[82]

$$
\frac{1}{\sigma^2}E\left\{\operatorname{tr}\left[\boldsymbol{W_A'}\boldsymbol{W_A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon'}\right]\right\} = \frac{1}{\sigma^2}\operatorname{tr}\left[\boldsymbol{W_A'}\boldsymbol{W_A}\right]\underbrace{E\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon'}\right]}_{\sigma^2\boldsymbol{I}} = \operatorname{tr}\left[\boldsymbol{W_A'}\boldsymbol{W_A}\right]. \tag{C.14}
$$

Combing these partial results leads to the corresponding entry in the information matrix

$$
\begin{aligned}
-E\left[-\operatorname{tr}\left(\boldsymbol{W_A}\boldsymbol{W_A}\right)-\frac{1}{\sigma^2}\left(\boldsymbol{W}\boldsymbol{y}\right)'\boldsymbol{W}\boldsymbol{y}\right] &= \operatorname{tr}\left[\left(\boldsymbol{W_A}+\boldsymbol{W_A}'\right)\boldsymbol{W_A}\right] \\
&\quad +\frac{1}{\sigma^2}\left(\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}.
\end{aligned}
$$

Substituting for $\boldsymbol{y}$ in the in the last entry in the second column and transforming the resulting expression in a similar manner as in Equation (C.14) leads to

$$
-E\left[-\frac{1}{\sigma^4}\boldsymbol{\varepsilon'}\boldsymbol{W}\boldsymbol{y}\right] = \frac{1}{\sigma 4}E\left[\operatorname{tr}\left(\boldsymbol{W_A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon'}\right)\right] = \frac{1}{\sigma^2}\operatorname{tr}\boldsymbol{W_A}.
$$

This entry is also identical to the second one in the third column in the information matrix, and the first entry in this column equals $\mathbf{0}$.[83] Noting that $E\left[\varepsilon'\varepsilon\right] = N\sigma^2$, the remaining entry on the diagonal reads

$$
-E\left[\frac{N}{2\sigma^4}-\frac{1}{\sigma^6}\boldsymbol{\varepsilon'}\boldsymbol{\varepsilon}\right] = -\frac{N}{2\sigma^4}+\frac{1}{\sigma^6}N\sigma^2 = \frac{N}{2\sigma^4}.
$$

Collecting the results for the individual entries derived above, leads to the following

[82] See also Anselin (1988b, 77) for the second equality in this derivation.

[83] This follows as this value is simply the transpose of the third entry in the first column.

information matrix

$$\boldsymbol{I}(\boldsymbol{\delta},\gamma,\sigma^2)=\begin{pmatrix} \frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} & \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)' & \boldsymbol{0} \\ \frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta} & \operatorname{tr}\left[\left(\boldsymbol{W_A}+\boldsymbol{W_A}'\right)\boldsymbol{W_A}\right]+\frac{1}{\sigma^2}\left(\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta} & \frac{1}{\sigma^2}\operatorname{tr}\boldsymbol{W_A} \\ \boldsymbol{0} & \frac{1}{\sigma^2}\operatorname{tr}\boldsymbol{W_A} & \frac{N}{2\sigma^4} \end{pmatrix}.$$

Finally, the asymptotic variance-covariance matrix, $\boldsymbol{V}(\boldsymbol{\delta},\gamma,\sigma^2)$, on which the hypotheses tests will be based, is then given by the inverse of the information matrix, i.e. $\boldsymbol{V}(\boldsymbol{\delta},\gamma,\sigma^2)=\boldsymbol{I}(\boldsymbol{\delta},\gamma,\sigma^2)^{-1}$.

D Additional Estimation Results

This appendix provides estimation results from two additional analyses. The estimation results in the first section demonstrate that the omission of the state of Delaware is crucial for the results regarding the significance of the estimate of the investment rate in physical capital divided by the effective depreciation rate. Next, in Section D.2, the time horizon of the analysis is extended to cover the period 1990-2007, thereby ignoring the warning by the Bureau of Economic Analysis mentioned in Footnote 43 of Section 5.1 about appending the data series for the dependent variable.

D.1 Results – Benchmark Sample not Omitting Delaware

Table D.1: Estimation Results for Three Different Models for the Baseline Sample plus the State of Delaware and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Model	Solow (1956)	Howitt (2000)	Ertur and Koch (2007)			Ertur and Koch (2011)		
Interaction matrix			$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Constant	11.337	11.755	10.931	10.969	10.949	10.933	11.041	11.011
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.179	0.231	0.169	0.148	0.117	0.155	0.150	0.119
	(0.233)	(0.119)	(0.188)	(0.288)	(0.406)	(0.242)	(0.301)	(0.419)
$\ln s_{A,i}$	—	0.072	—	—	—	−0.013	0.001	0.001
		(0.033)				(0.745)	(0.979)	(0.984)
$\ln n_i$	—	0.280	—	—	—	0.018	0.013	0.011
		(0.406)				(0.547)	(0.680)	(0.717)
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	—	—	−1.589	−0.232	0.061	−1.716	−0.258	0.034
			(0.101)	(0.608)	(0.920)	(0.084)	(0.572)	(0.956)
γ	—	—	0.149	0.129	0.135	0.159	0.126	0.133
			(0.000)	(0.001)	(0.001)	(0.003)	(0.021)	(0.014)
AIC	−3.661	−3.691	−3.869	−3.814	−3.807	−3.796	−3.736	−3.727
BIC	−3.583	−3.535	−3.498	−3.659	−3.651	−3.562	−3.502	−3.493
Number of observations	47	47	47	47	47	47	47	47

Note: p-values are given in parentheses.

Table D.2: Estimation Results for the Direct, Indirect and Total Impacts in the Multi-Region Schumpeterian Model for the Baseline Sample plus the State of Delaware and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Interaction matrix	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Direct impacts:			
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.155	0.150	0.119
	(0.248)	(0.306)	(0.422)
$\ln s_{A,i}$	−0.013	0.001	0.001
	(0.745)	(0.980)	(0.983)
$\ln n_i$	0.018	0.013	0.011
	(0.551)	(0.682)	(0.717)
Indirect impacts:			
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	0.027	0.019	0.015
	(0.321)	(0.421)	(0.537)
$\boldsymbol{W} \ln s_{A,j}$	−0.005	−0.002	−0.002
	(0.628)	(0.797)	(0.803)
$\boldsymbol{W} \ln n_j$	0.003	0.001	0.001
	(0.616)	(0.782)	(0.811)
Total impacts:			
$\frac{\ln s_{K,i}}{\ln(n_i+0.02+\delta_i)} + \boldsymbol{W} \frac{\ln s_{K,j}}{\ln(n_j+0.02+\delta_j)}$	0.183	0.169	0.134
	(0.248)	(0.308)	(0.428)
$\ln s_{A,i} + \boldsymbol{W} \ln s_{A,j}$	−0.017	−0.001	−0.001
	(0.718)	(0.987)	(0.984)
$\ln n_i + \boldsymbol{W} \ln n_j$	0.021	0.014	0.012
	(0.556)	(0.691)	(0.729)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

D.2 Data and Estimation Results – Baseline Sample (Extended Time Horizon: 1990-2007)

This appendix provides a brief description on how the variables have been constructed for the case when the sample period is extended to include the years 1990-1996 as well. After providing summary statistics in Table D.3, the results from the estimation of the nested models are shown in Table D.4. Estimates of the impacts are given in Table D.5.

Table D.3: Summary Statistics – Baseline Sample (Extended Time Horizon: 1990-2007).

Variable	Mean	Median	Standard deviation	Minimum	Maximum
y_i	85,012.07	80,896.36	13,921.60	66,616.49	123,281.63
$s_{K,i}$	0.079	0.075	0.014	0.061	0.127
n_i	0.016	0.014	0.009	0.002	0.043
δ_i	0.048	0.047	0.001	0.048	0.051
$n_i + g^w + \delta_i$	0.084	0.082	0.009	0.069	0.109
$s_{A,i}$	0.021	0.019	0.015	0.004	0.075
H_i	0.237	0.228	0.043	0.144	0.328
$\frac{s_{K,i}}{n_i+g^w+\delta_i}$	0.947	0.917	0.126	0.776	1.398
$\boldsymbol{W_1 s_K}$	0.219	0.214	0.044	0.130	0.332
$\boldsymbol{W_2 s_K}$	0.226	0.220	0.054	0.130	0.390
$\boldsymbol{W_3 s_K}$	0.224	0.219	0.050	0.128	0.345
$\boldsymbol{W_1 y}$	19,617.38	18,376.66	4,434.07	11,868.55	30,909.47
$\boldsymbol{W_2 y}$	19,825.09	18,874.88	5,152.72	11,411.44	40,053.49
$\boldsymbol{W_3 y}$	20,114.12	18,874.88	5,144.09	12,313.56	35,559.93

Note: The given values are the original values (i.e. not in logs) for the benchmark sample of 47 states and the period 1990-2007 with y_i the income per worker in 2007.

Even though the dependent variable is still real per worker income in 2007, values for the earlier years are needed to calculate the average real investment rate in physical capital, as Yamarik (2013) only provides values for gross real investment in physical capital. The data series on nominal gross state product for the years 1990-1996 from the Bureau of Economic Analysis' regional accounts data (BEA, 2015b) based on SIC, has been transformed as described in Section 5.1 into real 2000 dollars and then appended to the series for the years 1997-2007 based on NAICS.

Another complication arose in the construction of this data set, as the OECD only provides annual values for the R&D investment rate from 1997 onwards and additionally for the years 1991, 1993, and 1995 (OECD, 2015). Hence, the values for the years 1992 and 1994 have been interpolated by taking the average of the previous and subsequent years' value before calculating the average over the values from 1992 to 2007 to obtain the variable $s_{A,i}$.

As can be seen from Table D.3, no negative values for the employment growth rate occurred in this sample so that the values for all observations can be transformed into logs without any problems.

Finally, note that even though the dependent variable has not changed and the neighborhood relations and geographic distances between states are identical to the ones for the sample in the main text, this is not the case for the spatial lags, as these include a measure for the human capital stock.

Table D.4: Estimation Results for Three Different Models for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1990-2007.

Model	Solow (1956)	Howitt (2000)	Ertur and Koch (2007)			Ertur and Koch (2011)		
Interaction matrix			$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Constant	11.379	11.374	11.119	11.050	11.052	11.093	10.863	10.869
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.643	0.637	0.544	0.543	0.509	0.566	0.527	0.504
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
$\ln s_{A,i}$	—	0.044	—	—	—	0.020	−0.008	−0.007
		(0.134)				(0.505)	(0.809)	(0.826)
$\ln n_i$	—	−0.042	—	—	—	−0.042	−0.039	−0.040
		(0.180)				(0.153)	(0.182)	(0.161)
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	—	—	0.097	−0.489	−0.019	−0.041	−0.295	0.063
			(0.937)	(0.443)	(0.982)	(0.973)	(0.648)	(0.940)
γ	—	—	0.098	0.118	0.118	0.070	0.114	0.112
			(0.014)	(0.001)	(0.001)	(0.116)	(0.017)	(0.017)
AIC	−4.003	−4.056	−4.035	−4.141	−4.130	−4.019	−4.086	−4.082
BIC	−3.925	−3.898	−3.877	−3.983	−3.972	−3.783	−3.850	−3.846
Number of observations	47	47	47	47	47	47	47	47

Note: p-values are given in parentheses.

Table D.5: Estimation Results for the Direct, Indirect and Total Impacts in the Multi-Region Schumpeterian Model for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1990-2007.

Interaction matrix	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Direct impacts:			
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.565	0.527	0.504
	(0.000)	(0.001)	(0.002)
$\ln s_{A,i}$	0.020	-0.008	-0.007
	(0.508)	(0.812)	(0.824)
$\ln n_i$	-0.042	-0.039	-0.040
	(0.160)	(0.189)	(0.168)
Indirect impacts:			
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	0.042	0.066	0.062
	(0.155)	(0.049)	(0.053)
$\boldsymbol{W} \ln s_{A,j}$	0.001	-0.002	-0.002
	(0.765)	(0.667)	(0.675)
$\boldsymbol{W} \ln n_j$	-0.003	-0.005	0.006
	(0.358)	(0.294)	(0.957)
Total impacts:			
$\frac{\ln s_{K,i}}{\ln(n_i+0.02+\delta_i)} + \boldsymbol{W} \frac{\ln s_{K,j}}{\ln(n_j+0.02+\delta_j)}$	0.607	0.593	0.566
	(0.000)	(0.001)	(0.002)
$\ln s_{A,i} + \boldsymbol{W} \ln s_{A,j}$	0.021	-0.010	-0.001
	(0.518)	(0.788)	(0.801)
$\ln n_i + \boldsymbol{W} \ln n_j$	-0.045	-0.044	-0.045
	(0.161)	(0.189)	(0.169)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

E List of States Included in the Empirical Analyses

This appendix lists the states that are included in the different empirical analyses and also provides a correspondence with the state abbreviations used in Figure 1.

Table E.6: Alphabetical List of the 48 US States plus the District of Columbia.

State	Code	State	Code
Alabama	AL	Nebraska	NE
Arizona	AZ	Nevada	NV
Arkansas	AR	New Hampshire	NH
California	CA	New Jersey	NJ
Colorado	CO	New Mexico	NM
Connecticut	CT	New York	NY
Delaware	DE	North Carolina	NC
District of Columbia	DC	North Dakota	ND
Florida	FL	Ohio	OH
Georgia	GA	Oklahoma	OK
Idaho	ID	Oregon	OR
Illinois	IL	Pennsylvania	PA
Indiana	IN	Rhode Island	RI
Iowa	IA	South Carolina	SC
Kansas	KS	South Dakota	SD
Kentucky	KY	Tennessee	TN
Louisiana	LA	Texas	TX
Maine	ME	Utah	UT
Maryland	MD	Vermont	VT
Massachusetts	MA	Virginia	VA
Michigan	MI	Washington	WA
Minnesota	MN	West Virginia	WV
Mississippi	MS	Wisconsin	WI
Missouri	MO	Wyoming	WY
Montana	MT		

References

Absil, P.-A., R. Mahony, and R. Sepulchre (2008). *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton.

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*, Princeton University Press, Princeton.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

Aghion, P., U. Akcigit, and P. Howitt (2014). "What Do We Learn From Schumpeterian Growth Theory?" in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 2B, Chapter 1, Elsevier B.V., Oxford, 515–563.

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

——— (1998). *Endogenous Growth Theory*, MIT Press, Cambridge.

Akaike, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle," in B. N. Petrov and F. Cszaki (editors), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267–281.
Reprinted in: S. Kotz and N. L. Johnson (editors) (1992), *Breakthroughs in Statistics Volume 1. Foundations and Basic Theory*, Springer Series in Statistics, Perspectives in Statistics, Springer, Berlin, 610-624.

Anselin, L. (1988a). "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity," *Geographical Analysis*, 20(1), 1–17.

——— (1988b). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon (1996). "Simple diagnostic tests for spatial dependence," *Regional Science and Urban Economics*, 26, 77–104.

Arrow, K. J. (1962). "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29, 155–173.

Barro, R. J. and X. Sala-i-Martin (1992). "Convergence," *Journal of Political Economy*, 100(2), 223–251.

——— (2004). *Economic Growth*, 2nd Edition, MIT Press, Cambridge.

Basile, R. (2014). "Regional productivity growth in Europe: a Schumpeterian perspective," Geocomplexity Discussion Paper No. 1.

Bode, E., P. Nunnenkamp, and A. Waldkirch (2012). "Spatial effects of foreign direct investment in US states," *Canadian Journal of Economics*, 45(1), 16–40.

Bureau of Economic Analysis (2015a). *National Economic Accounts.*
URL: *http://www.bea.gov/national/index.htm, accessed: 29 May, 2015.*

——— (2015b). *Regional Economic Accounts.*
URL: *http://www.bea.gov/regional/index.htm, accessed: 22 June, 2015.*

Bureau of Labor Statistics (2015). *Current Employment Statistics.*
URL: *http://www.bls.gov/sae/, accessed: 22 June, 2015.*

Caballero, R. J. and A. B. Jaffe (1993). "How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in O. Blanchard and S. Fischer (editors), *NBER Macroeconomis Annual*, Volume 8, Chapter 1, MIT Press, Massachusetts, 15–86.

Cheney, W. and D. Kincaid (2008). *Numerical Mathematics and Computing*, 6th Edition, Thomson Higher Education, Belmont.

Cliff, A. and K. Ord (1972). "Testing for Spatial Autocorrelation Among Regression Residuals," *Geographical Analysis*, 4(3), 267–284.

Dalgaard, C.-J. and H. Strulik (2013). "The history augmented Solow model," *European Economic Review*, 63, 134–149.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*, Oxford University Press, Oxford.

Dowdy, S., S. Wearden, and D. Chilko (2004). *Statistics for Research*, 3rd Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken.

Durlauf, S. N. and D. T. Quah (1999). "The New Empirics of Economic Growth," in J. Taylor and M. Woodford (editors), *Handbook of Macroeconomics*, Volume 1A, Chapter 4, Elsevier Science B.V., Oxford, 235–308.

Eaton, J. and S. Kortum (1996). "Trade in ideas Patenting and productivity in the OECD," *Journal of International Economics*, 40, 251–278.

——— (1999). "International Technology Diffusion. Theory and Measurement," *International Economic Review*, 40(3), 537–570.

Elhorst, P. J. (2010). "Applied Spatial Econometrics: Raising the Bar," *Spatial Economic Analysis*, 5(1), 9–28.

Ertur, C. and W. Koch (2007). "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence," *Journal of Applied Econometrics*, 22(6), 1033–1062.

——— (2011). "A contribution to the theory and empirics of Schumpeterian growth with worldwide interactions," *Journal of Economic Growth*, 16, 215–255.

Fischer, M. M. (2011). "A spatial Mankiw-Romer-Weil model: theory and evidence," *The Annals of Regional Science*, 47, 419–436.

FISCHER, M. M., M. BARTKOWSKA, A. RIEDL, S. SARDADVAR, AND A. KUNNERT (2010). "The Impact of Human Capital on Regional Labor Productivity in Europe," in M. M. Fischer and A. Getis (editors), *Handbook of Applied Spatial Analysis – Software Tools, Methods and Applications*, Chapter E.1, Springer, Berlin, 585–597.

FLORAX, R. J., H. FOLMER, AND S. J. REY (2003). "Specification searches in spatial econometrics - the relevance of Hendry's methodology," *Regional Science and Urban Economics*, 33, 557–579.

GAROFALO, G. A. AND S. YAMARIK (2002). "Regional Convergence: Evidence from a New State-by-State Capital Stock Series," *The Review of Economics and Statistics*, 84(2), 316–323.

GERSCHENKRON, A. (1962). *Economic Backwardness in Historical Perspective: A Book of Essays*, The Belknap Press of Harvard University Press, Cambridge.

GERSCHGORIN, S. (1931). "Über die Abgrenzung der Eigenwerte einer Matrix," *Bulletin de L'Académie des Sciences de L'URSS*, 749–754.

GILFILLAN, S. C. (1935a). *Inventing the Ship*, Follett, Chicago.

——— (1935b). *The Sociology of Invention*, Follett, Chicago.

GREENE, W. H. (2003). *Econometric Analysis*, 5th Edition, Pearson Education Ltd., Upper Saddle River.

HALBERSTAM, D. AND R. M. HILLS, JR. (2001). "State Autonomy in Germany and the United States," *Annals of the American Academy of Political and Social Science*, 574, 173–184.

HANUSHEK, E. A., J. RUHOSE, AND L. WOESSMANN (2015). "Human Capital Quality and Aggregate Income Differences: Development Accounting for U.S. States," Working Paper 21295, National Bureau of Economic Research.

HOLTZ-EAKIN, D. (1993). "Solow and the States: Capital Accumulation, Productivity, and Economic Growth," *National Tax Journal*, 46(4), 425–439.

HOWITT, P. (1999). "Steady Endogenous Growth with Population and R & D Inputs Growing," *Journal of Political Economy*, 107(4), 715–730.

——— (2000). "Endogenous Growth and Cross-country differences," *American Economic Review*, 90, 829–846.

ISLAM, N. (1995). "Growth Empirics: A Panel Data Approach," *The Quarterly Journal of Economics*, 110(4), 1127–1170.

ISPHORDING, I. E. AND S. OTTEN (2013). "The Costs of Babylon - Linguistic Distance in Applied Economics," *Review of International Economics*, 21(2), 354–369.

JONES, C. I. (1995a). "R & D-Based Models of Economic Growth," *The Journal of Political Economy*, 103(4), 759–784.

——— (1995b). "Time Series Tests of Endogenous Growth Models," *The Quarterly Journal of Economics*, 110(2), 495–525.

Keilbach, M. (2000). *Spatial Knowledge Spillovers and the Dynamics of Agglomeration and Regional Growth*, Physica-Verlag, Heidelberg.

Kelejian, H. H. and I. R. Prucha (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of Econometrics*, 157, 53–67.

Keller, K. R. I. and P. Poutvaara (2005). "Growth in OECD countries and elsewhere: how much do education and R&D explain?" *Economics Bulletin*, 15(16), 1–11.

Keller, W. (2002). "Geographic localization of international technology diffusion," *American Economic Review*, 92(1), 120–142.

Knowles, S. and P. D. Owen (1995). "Health capital and cross-country variation in income per capita in the Mankiw-Romer-Weil model," *Economics Letters*, 48, 99–106.

Krugman, P. (1991). "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

Le Gallo, J. (2014). "Cross-Section Spatial Regression Models," in M. M. Fischer and P. Nijkamp (editors), *Handbook of Regional Science*, Chapter 76, Springer-Verlag, Berlin, 1511–1533.

Lee, L. (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72(6), 1899–1925.

LeSage, J. and R. K. Pace (2009). *Introduction to Spatial Econometrics*, Chapman and Hall/CRC, Boca Raton.

LeSage, J. P. and K. R. Pace (2014). "The Biggest Myth in Spatial Econometrics," *Econometrics*, 2, 217–249.

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107, 407–437.

Melitz, J. and F. Toubal (2014). "Native language, spoken language, translation and trade," *Journal of International Economics*, 93, 351–363.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia.

National Science Board (2014). *Science and Engineering Indicators 2014*, National Science Foundation, Arlington.

Nelson, R. R. and E. S. Phelps (1966). "Investment in Humans, Technological Diffusion and Economic Growth," *American Economic Review*, 61(1/2), 69–75.

Nonneman, W. and P. Vanhoudt (1996). "A Further Augmentation of the Solow Model and the Empirics of Economic Growth for OECD Countries," *The Quarterly Journal of Economics*, 111(3), 943–953.

Ord, K. (1975). "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70(349), 120–126.

Organisation for Economic Co-operation and Development (2010). *OECD Science, Technology and Industry Outlook*, OECD Publishing.

——— (2014). *OECD Science, Technology and Industry Outlook*, OECD Publishing.

——— (2015). *OECD Regional Database.*
URL: *https://www.oecd.org/gov/regional-policy/regionalstatisticsandindicators.htm, accessed: 12 June, 2015.*

Pace, R. K. and R. Barry (1997). "Quick Computation of Spatial Autoregressive Estimators," *Geographical Analysis*, 29(3), 232–247.

Peri, G. (2012). "The Effect of Immigration on Productivity: Evidence From U.S. States," *The Review of Economis and Statistics*, 94(1), 348–358.

Ram, R. (2007). "IQ and economic growth: Further augmentation of Mankiw-Romer-Weil model," *Economis Letters*, 94, 7–11.

Rey, S. J. and B. D. Montouri (1999). "US Regional Income Convergence: A Spatial Econometric Perspective," *Regional Studies*, 33(2), 143–156.

Romer, P. M. (1990). "Endogenous Technological Change," *The Journal of Political Economy*, 98(5), S71–S102.

Ross, S. M. (2010). *Introduction to Probability Models*, 10th Edition, Elsevier, Amsterdam.

Samuelson, P. A. (1954). "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments," *The Economic Journal*, 64(254), 264–289.

Sarel, M. (1996). "Nonlinear Effects of Inflation on Economic Growth," *IMF Staff Papers*, 43(1), 199–215.

Schwarz, G. (1978). "Estimating the Dimension of a Model," *The Annals of Statistics*, 6(2), 461–464.

Solow, R. M. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70(1), 65–94.

Spolaore, E. and R. Wacziarg (2009). "The Diffusion of Development," *The Quarterly Journal of Economics*, 124(2), 469–529.

Swan, T. W. (1956). "Economic Growth and Capital Accumulation," *The Economic Record*, 32(2), 334–361.

Sydsæter, K., P. Hammond, A. Seierstad, and A. Strøm (2008). *Further Mathematics for Economic Analysis*, 2nd Edition, Prentice Hall, Harlow.

The Economist (2013). "The Missing $20 Trillion - How to stop companies and people dodging tax, in Delaware as well as Grand Cayman," February 16.
URL: *http://www.economist.com/news/leaders/21571873-how-stop-companies-and-people-dodging-tax-delaware-well-grand-cayman-missing-20, accessed: 11 August, 2015.*

Thomas Jr., G. B. (2005). *Thomas' Calculus*, 11th Edition, Addison-Wesley, Boston.

United States Census Bureau (2015). *Current Population Survey.*
URL: *http://www.census.gov/hhes/socdemo/education/data/cps/index.html, accessed: 22 June, 2015.*

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd Edition, W.W. Norton & Company, New York.

Verbeek, M. (2004). *A Guide to Modern Econometrics*, 2nd Edition, John Wiley & Sons, Ltd, Chichester.

von Thünen, J. H. (1826). *Der isolirte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, oder Untersuchungen über den Einfluß, den die Getreidepreise, der Reichthum des Bodens und die Abgaben auf den Ackerbau ausüben*, Friedrich Perthes, Hamburg.
URL: *http://www.deutschestextarchiv.de/book/show/thuenen_staat_1826, accessed: 29 July, 2015.*

Wälde, K. (2011). *Applied Intertemporal Optimization, Edition 1.1*, Mainz University Gutenberg Press.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*, 5th Edition, South-Western, Mason.

Yaari, M. E. (1965). "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *The Review of Economic Studies*, 32(2), 137–150.

Yamarik, S. (2006). "Solow and the States: New Evidence," *Regional Studies*, 40(6), 571–582.

——— (2011). "Human capital and state-level economic growth: what is the contribution of schooling?" *The Annals of Regional Science*, 47, 195–211.

——— (2013). "State-Level Capital and Investment: Updates and Implications," *Contemporary Economic Policy*, 31(1), 62–72.

Young, A. (1998). "Growth without Scale Effects," *Journal of Political Economy*, 106(1), 41–63.

Working Paper Series in Economics

recent issues

No. 75 *Tim Deeken:* Schumpeterian growth with technological interdependence: An application to US states, November 2015

No. 74 *Tim Deeken:* Knowledge spillovers: On the impact of genetic distance and data revisions, November 2015

No. 73 *Christian Feige and Karl-Martin Ehrhart:* Voting and transfer payments in a threshold public goods game, November 2015

No. 72 *Steffen Huck, Nora Szech, Lukas M. Wenner:* More effort with less pay: On information avoidance, belief design and performance, September 2015

No. 71 *Florian Kreuchauff and Vladimir Korzinov:* A patent search strategy based on machine learning for the emerging field of service robotics, August 2015

No. 70 *Christian Feige:* Success rates in simplified public goods games - a theoretical model, June 2015

No. 69 *Markus Fels:* Mental accounting, access motives, and overinsurance, May 2015

No. 68 *Ingrid Ott and Susanne Soretz:* Green attitude and economic growth, May 2015

No. 67 *Nikolaus Schweizer and Nora Szech:* Revenues and welfare in auctions with information release, April 2015

No. 66 *Andranik Tangian:* Decision making in politics and economics: 6. Empirically constructing the German political spectrum, April 2015

No. 65 *Daniel Hoang and Martin Ruckes:* The effects of disclosure policy on risk management incentives and market entry, November 2014