PALKANSAAJIEN TUTKIMUSLAITOS
LÖNTAGARNAS FORSKNINGSINSTITUT
LABOUR INSTITUTE FOR ECONOMIC RESEARCH

TYÖPAPEREITA | WORKING PAPERS 320

The Role of Natural Resources and Geography for Productivity in Developed Countries*

Eero Lehto**
Sakari Lähdemäki***
Eero Mäkynen****

* We gratefully acknowledge the funding provided by Palkansaajasäätiö.

PALKANSAAJA-SÄÄTIÖ

** Labour Institute for Economic Research (eero.lehto@labour.fi).
*** Labour Institute for Economic Research (sakari.lahdemaki@labour.fi).
**** University of Turku (eeanma@utu.fi).

Abstract

This paper considers how natural resources and geography impact labour productivity either directly or through R&D-intensity, education level and product structure. The country set considered includes 42 developed and developing countries. By geography, we mean population size, internal density and geographical centrality with respect to activity outside a country's borders. According to our results, of natural resources forest area and gas and oil reserves have mostly a negative total effect on labour productivity. Agricultural area and minerals, instead, support productivity. Of geographical variables internal density raises R&D-intensity and via it productivity. Geographical centrality and population size, instead, negatively impacts productivity.

JEL codes: O13, O40, O47, O57, Q00, R120

Keywords: productivity, natural resources, geography, R&D-intensity, education, consumer goods share of manufacturing

Tiivistelmä

Tämä tutkimus tarkastelee luonnonvarojen ja maantieteen vaikutusta työn tuottavuuteen ja asukaslukuun suhteutettuun BKT:hen eri maissa. Vaikutukset ovat joko suoria tai ne välittyvät T&K-intensiteetin, koulutuksen tason ja/tai teollisuuden tuotantorakenteen kautta. Tutkimus tarkastelee 42 kehittynyttä ja kehittyvää maata. Tutkimuksen maatiedemuuttujat ovat väestötiheys, maantieteellinen keskeisyys ja väestön koko. Tulosten mukaan maatalousmaan osuuden, metsämaan osuuden sekä kaasu- ja öljyvarojen kokonaisvaikutus tuottavuuteen on keskimäärin negatiivinen. Mineraalivaroilla, joita mitattiin mineraalien tuotannolla, ja maataloudella on positiivinen vaikutus. Maantiedemuuttujista väestötiheys lisää T&K-investointeja ja nostaa sitä kautta tuottavuutta. Tämä vaikutus syntyy siitä huolimatta, että koulutustaso reagoi väestötiheyteen negatiivisesti. Maan keskeinen sijainti, sen rajojen läheisyydessä olevan taloudellisen aktiviteetin perusteella, vaikuttaa negatiivisesti tuottavuuteen. Vastoin odotuksia väestön koon kokonaisvaikutus tuottavuuteen on negatiivinen useimmissa estimoiduissa malleissa. Vaikutus syntyy, kun väestön koko supistaa T&K-investointeja ja alentaa koulutustasoa.

JEL koodit: O13, O40, O47, O57, Q00, R120

Avainsanat: tuottavuus, luonnonvarat, maantiede, T&K-intensiteetti, koulutus, kulutustavaroiden osuus teollisuuden tuotannosta

Palkansaajien tutkimuslaitos
Labour Institute for Economic Research
Pitkänsillanranta 3 A
00530 Helsinki

Työpapereita | Working Papers 320
ISBN 978-952-209-172-7 (pdf)
ISSN 1795-1801 (pdf)

www.labour.fi

Helsinki 2018

1. Introduction

This paper considers how natural resources and geography impact the GDP and productivity of a country set that mostly includes developed countries. By geography, we mean the average population and density of the country and its centrality with respect to activity outside its borders.

In the field of economic geography, it is recognized that the interaction of two forces – namely, the tendency to utilize comparative advantages and the tendency to utilize increasing returns and, in consequence, to specialize in certain activities – shapes the economic structures of different countries (see Helpman and Krugman, 1985 and Krugman 2009).[1] As noted by Deardorff (2001), in the post-war era, many developing countries became labour-intensive manufacturers, and the most advanced developed countries concentrated on skill-intensive stages of production. Specialization of this type does not necessarily generate differences in the central factors of production, but it creates an opportunity for one country to differentiate itself from another in terms of production structure. Then, some initial state in natural resources or geography can push a country's economy in a specific direction.

Relying on the economic model based on increasing returns and product differentiation, Krugman (1980) introduced the so-called home market effect (HME) hypothesis, according to which countries tend to export those goods for which they have relatively large domestic markets.[2] The empirical relevance of this hypothesis has been tested many times in the economic literature. David and Weinstein (2003) discovered that in the presence of trade costs, the home market effect is characterized by a major part of the manufacturing production in OECD countries. The rest of the manufacturing production is governed by simple comparative advantage. Hanson and Xiang (2004) found evidence consistent with the home-market effect for many typical heavy intermediate goods industries. Behrens et al. (2004) discovered that such industries as industrial chemicals, fabricated metals and metal products, electrical on non-electrical machinery and transport equipment obey the home market effect. These later results confirm the preconception that HME concerns the industries with large fixed costs, differentiated product and remarkable transport costs.

Our study does not explicitly test the relevance of demand-side HME hypothesis. In the growth context of our study, the geographic impacts may also resemble factor endowments in the sense that the availability of knowledge capital may impact the industrial structure and economy in many ways. However, we do not exclude the possible existence of the HME phenomenon. In our model, in which the manufacturing production is divided in consumer goods production and other (intermediate and investments goods)

[1] The empirical findings of Helpman (1987) show that increasing returns, monopolistic competition and product differentiation induce seemingly similar countries to specialize in the production of goods that belong to the same industry as the goods produced by their trading partners.

[2] See also Helpman (1981).

production, HME is not necessarily seen.[3] Presumably, insofar as any impact is found, the size of a country would increase the share of intermediate and investment goods production because increasing returns and large transport costs are typical of these industries. In the empirical analysis of this study, we also explore the impacts of the geographic centrality. In this respect, our study touches the research that tests the existence of HME, also taking into account the proximity to other countries, as in Behrens et al. (2004).

Motivated by the pioneering study of Romer (1990), our study of labour productivity and GDP per capita emphasizes the importance of intellectual capital in forming the level of productivity and economic structure. Thus, rather than labour and capital intensity in the production, it is the knowledge intensity that is seen to impact the level of productivity. In this respect, we rely strongly on the factor endowments. In the country set analysed, differences in the relative abundance of traditional the production factors of labour and capital do not seem to create comparative advantages, whereas differences in the relative abundance of labour skills and available technology do create such advantages. In this respect, our approach is close to Bernard et al. (2007), who consider the impacts of trade liberalisation in the frameworks where firms and countries differ from each other in respect to skill intensiveness.[4]

In this study, “initial state” refers to states in geography and natural resources. We test the impact of these initial states on the relative abundance of the central forces of production –, i.e., R&D and education – and on production structure and, through these factors, on productivity. We also try to identify the mechanism of these impacts. Specifically, are they due exclusively to market forces or is policy intervention also involved?

Evidence in the literature regarding the effects of natural resources is mixed.[5] According to Sachs and Warner (1999 and 2001), an abundance of natural resources has a negative impact. If a country acquires the greatest advantage in terms of natural resources, the resulting impairment of cost competitiveness would outweigh the impact on GDP. In contrast, Torres and Afonso (2008) find that natural resources have a positive impact on economic growth. Boschini et al. (2007 and 2013) emphasize that natural resources are not alike, having different degrees of “technical appropriability”, and that the negative association of natural resource intensity and economic growth can be reversed if the institutional quality is high enough. We take this variability into account by including several natural resource variables in our econometric model.

The growth literature has considered various impacts of geography in different ways. For example, Gallup et al. (1999) try to capture the impacts of pure geography (i.e., terrain, climate and geographical location), whereas others take into account population concentration by considering the dimensions of population

[3] The evidence for HME would be found more easily with more disaggregated data, as shown by Davis and Weinstein (1996).
[4] According to their results, the countries and firms with abundant skills benefit from the net of trade liberalisation in terms of changes in productivity and in the number of jobs.
[5] According to Havranek et al., 2016, only 20 percent of these studies found this impact positive.

density or urbanisation, which moreover can be considered an outcome of economic development or a phenomenon closely linked to such development rather than its driver.[6]

Many studies argue that various forms of agglomeration – more or less related to density – favour economic growth. For example, the results obtained by Delgado et al. (2014) indicate that overlapping clusters favour economic growth and that the diminishing returns caused by possible congestion have no remarkable influence. Research on knowledge spillovers shows that R&D or patenting activities have a positive spillover effect on other firms if they are sufficiently proximate to one another[7]. The low cost of transportation is another feature of market proximity (or density) that has a positive growth effect.[8]

Studies that analyse the impacts of pure geography (or at least features that are more invariant than urbanisation) focus mainly on the north-south axis and the dilemma of underdevelopment. Gallup et al. (1999) analyse GDP growth using substantial country data regarding all types of countries and discover that coastal density has a strong positive impact on growth. They also show that population density is determined by pure geography (i.e., climate, terrain, diseases and distances). Moreover, Gallup et al. (1999) find that variables such as the quality of public institutions and trade openness account for much of the explanatory power of geographic variables. Nonetheless, instead of concluding that institutions matter most, they determine that favourable geography has a remarkable impact on economic growth.

In empirical growth studies, geography is often viewed as an invariant that acts through certain endogenous intermediators. Studies that consider the underdevelopment of tropical countries and the growth dilemma along the south-north axis often see institutions as a variable intermediating force with a positive impact on growth (see Acemoglu et al. (2001) and (2002), Easterly and Levine (2002), Rodrik et al. (2004). Typical geographic variables in these studies are tropical climate, distance from the equator or the sea, and morbidity. In the abovementioned growth studies, institutional variables (which in Acemoglu et al. (2001 and 2002) include institutions that resulted from countries' colonial pasts) capture the entire direct effect of pure geography on economic growth.

Institutions include property rights (Vieira et al. 2012), the concentration of power and the colonial past (Acemoglu et al., 2002), and factors that describe the state of the public sector, democracy and trade openness (see Easterly and Levine (2003) and Rodrik et al. (2004)). The basic dilemma in these studies is the

[6] For example, Hendersson (2003) views urbanisation as a sign of a structural shift within society that is related to development. Hendersson (2003) and later Brülhart and Sbergami (2009) obtain results indicating that population concentration in large cities initially favours economic growth but that urbanization is no longer related to economic growth at higher levels of development.
[7] See Jaffe *et al*., (1993), Maurseth and Verspagen, (2002), Keller, (2002), Grünfeld (2002) and Lehto, (2007).
[8] In Gallup et al. (1999), the relevant variable was shipping costs.

endogeneity of institutions. Even when an instrumental variable approach is taken (as it was in Rodrik et al., 2004), the question about whether institutions are a cause or a consequence remains problematic.

According to Glaeser et al. (2004), institutions are an outcome of economic development rather than a cause. In their analysis, school years in the first year of the studied time interval seem to capture the growth impact of institutions. The authors ultimately conclude that human capital is a more basic source of economic growth than institutions are.

The purpose of this study is not only to consider the overall impacts natural resources and geography on the productivity in developed (and some developing) countries but also and more specifically to examine the anatomy of these impacts. There are many reasons to believe that these impacts are two-fold. They may have either adverse or positive effects, as the economic literature on natural resources has discovered. We hypothesize that the same applies to different dimensions of geography, as well.

Considering this duality, we assess whether the impact is direct or total so that it occurs through one or more intermediate variables, such as production structure (the share of consumer goods production in manufacturing), R&D intensity and adult skill levels (education). Education and R&D are selected as intermediator variables based on modern growth theory, wherein the accumulation of human capital (and/or knowledge capital) plays a central role. The share of consumer goods production is thought to reflect the quality requirements of manufacturing and, more generally, the service production linked to manufacturing. Simply put, we believe that quality standards are higher for intermediate and investment goods than for consumer goods because the former goods and related services are bought by other firms, whereas consumer goods are bought by consumers whose purchase habits are influenced not only by quality but also by other criteria, such as image. In addition, intermediate and investment goods production has higher transportation costs and exhibits increasing returns to scale more clearly than consumer goods production.[9]

By introducing different types of intermediate variables, we test the possible existence not only of negative "crowding out" effects of natural resources or geography but also of market and policy reactions that may be motivated by efforts to compensate for disadvantages that stem from natural resources and geography. The latter aspect distinguishes this study from the previous literature.

This study mainly considers the differences among developed countries. The geographic and natural resource variables are then chosen to account for different features of these countries. The geographic variables included in our study are population size, various measures of internal density, and centrality, which defines

[9] According to Hanson and Xiang (2004), (see Table 2 on page 34) the heavy industry – typical producers of intermediate goods - has relatively high transport goods. This could also apply to investment goods that presumably are more complex and vertically more specialized than consumer goods. Anyway, for goods that are produced in multiple stages transportation costs are relatively high as is suggested in World Trade Report (2008).

the closeness of the activity outside a country's borders. We also comment results regarding the impacts of urbanization that can, however, be also considered an outcome of a phenomenon linked closely to economic development rather than its driver. The natural resources considered are the share of agricultural area, the share of forest area, gas and oil reserves per capita and minerals production per capita.

We also tend to control the development stage of the countries considered. Recent research has shown that proximity to the technological frontier increases the profitability of investments in R&D and in higher education (see Acemoglu et al. 2006, Vandenbussche at al. 2006, and Lehto et al., 2010). These observations indicate that countries at different stages of development have different factor compositions in terms of human capital. The modern growth theory suggests that in advanced countries – which we mostly consider in this study – human capital and the other elements of knowledge capital are emphasized as production factors. In our study, this notion is tested in the context of other geographic variables. For example, the average distance (between randomly located economic agents) within the country measures agglomeration, at least to some extent. It is thus interesting to test whether the impact of this variable on R&D depends on the level of productivity.

In the empirical model of this study, the intermediating variables (R&D-intensity, production structure and education level) are considered endogenous and time varying regressors. The geographical and natural resources variables, instead, are time invariant and exogenous. The inclusion of both time-varying endogenous variables and time-invariant exogenous variables in our model causes a problem, which we address by estimating our panel data model using the Hausman-Taylor fixed effects estimator, which is designed to be applied in this type of situation. This procedure also gives an estimate for time-invariant variables and is efficient for both time-varying and time-invariant variables.[10]

The whole setting in which the results are realized either directly or indirectly via intermediate variables in the present form is new in the literature. Therefore, we are able to discover such impacts that cannot be found when only the total impacts are observed. In detail, according to the central results obtained, agricultural land area positively impacts productivity despite the fact that agricultural area has a tendency to contract R&D investments. The total impact of forest area and gas and oil reserves on productivity is slightly negative. However, this negativity is diluted by the fact that forest area increases the share of intermediate

[10] The difficulty of estimating coefficients for regressors that vary only slightly in time has recently drawn attention in the growth literature. Fixed effects estimation cannot identify coefficients for time-invariant variables. In contrast, the random effects model suffers from endogeneity bias. Accordingly, Barro (2015) in his study of convergence discusses the challenges of finding a statistically significant estimate for a slightly varying variable in a regression where country fixed effects are included. Barro (2015) ultimately uses panel OLS estimation and two-stage least-squares estimation with lagged values of the variables included as instruments.

goods and investment goods production and that gas and oil reserves support education. Minerals (which are measured in terms of their production) support productivity via their positive impact on R&D intensity.

In most models, population size has a negative total impact on productivity or GDP per capita. This is mainly because population size, by its direct impact, slightly lowers the education level and decreases R&D intensity. The latter effect, in absolute value, is larger the higher the country's initial level of productivity is. Population density (or the inverse of the distance within the country) and related externalities of agglomeration – by its direct impact and total impact - strongly raise a country's R&D intensity and therefore also the level of its productivity. It is remarkable that the direct reaction of education to the internal density goes in the opposite direction. The individual choices, the markets or the social and economic policy in their actions in education then tends to attenuate the disadvantages of low density. The country's geographic location – being central or remote – also has some unexpected impacts. The overall effect of centrality on productivity is negative in most models, because centrality lowers R&D intensity. The centrality increases the share of intermediate and investment goods production weakening its negative total effect on productivity. That remoteness increases R&D investments can be regarded as a kind of compensating reaction. All in all, that education and R&D intensity seem to counteract the disadvantages of sparse population and remote location via policy actions or market repercussions are novel results in the literature.

Our geographical and natural resources variables are time averages from years 1990-1994. The cross-sectional differences of these variables have been similar for at least several decades. Owing to this, these variables can be regarded as exogenous. The use of time averages from a period that precedes the estimation period deepens the exogenous nature of these variables. Note that the coefficients of geographical and natural resource variables are not identified from contemporary time variation which in many cases is not independent of the state of the endogenous variables of this study. Given our country set, data limitations set boundaries to the choice over the set of variables in the estimated regressions. Therefore, we can't completely avoid the vulnerabilities raising from possible omitted variable bias. Anyway, we interpret that the estimated coefficients of geographical and natural resources variables describe more or less causal impacts which are either direct or will be intermediated to productivity through the endogenous variables.

The rest of the paper is structured as follows. In section 2 we present the hypotheses we test and discuss the theoretical models (more detailed description in the appendixes) underlying these hypotheses. In section 3 we describe our dataset. The estimation method and the estimated model is presented in section 4 whereas the main results are reported in section 5. In section 6 we report result for some additional variables and robustness checks. Section 7 concludes.

2. Theoretical models and testable hypotheses

In appendixes (chapter A2.1), we introduced a Schumpeterian growth model that captures the growth impacts of various parameters reflecting natural advantages and disadvantages of a country's geography and natural resources. In the constructed model – in which the monopolist intermediate goods producer invests in innovations – the final goods production is divided in two sectors: the production of intermediate and investment goods and the production of consumer goods. These sectors differ from each other in their efficiencies to accomplish innovations. Furthermore, the model includes the supply and demand of natural resources. In another model (chapter A2.2.), we consider education decisions. In the introduced framework, the immigration and emigration of talents may dilute the social benefits of educational investments of densely populated and centrally located countries.

In the following, we present some empirically testable hypotheses which are derived from the above models and previous results of the literature.

- We believe that the density within a country creates agglomeration externalities and increases the intensity of R&D. In the simulations of model A2.1, this is taken into account by increasing the parameter λ.
- We believe that in the purchase of intermediate and investment goods, only the quality matters, whereas decisions to purchase consumer goods are also governed by other motives. Owing to this, technological standards are more demanding in the production of intermediate and investment goods than in the production of consumer goods. In our model framework, this leads to the situation in which costs to generate a unit value of innovation are lower in the former sector than in the latter sector. Owing to this – as we show with model simulations - R&D intensity tends to correlate positively with the share of intermediate and investment goods production.
- The increase in the abundance of natural resources tends to expand GDP and lower the intensity of R&D, as we show with model simulations. Only if innovating in intermediate and investment goods is exceptionally efficient would the increase in the supply of natural resources exclusively suitable for this sector raise R&D intensity in the whole economy. This is also explored in the empirical part of this study.
- The most difficult question is related to the impacts of the relevant market size. The realization of the so-called home market effect could mean that the share of intermediate and investment goods production is larger in a large country than in a small country. This would result in higher R&D intensity and productivity in the large country. By model simulation, we produce a corresponding result. Empirical verification is, however, impaired by the fact that market size is measured by population size and not by GDP and that manufacturing is divided only roughly into two sectors.

- According to the counter hypothesis to the above hypothesis, the home bias in consumer goods purchases and simultaneous scale economies matter most. This would grow the relative size of consumer goods production in large country and respectively lower the R&D intensity and productivity.
- The analysis in chapter A2.2 allows us to hypothesize that internal density and central location could diminish school investments and weaken individuals' educational efforts. This is also tested in the empirical analysis.

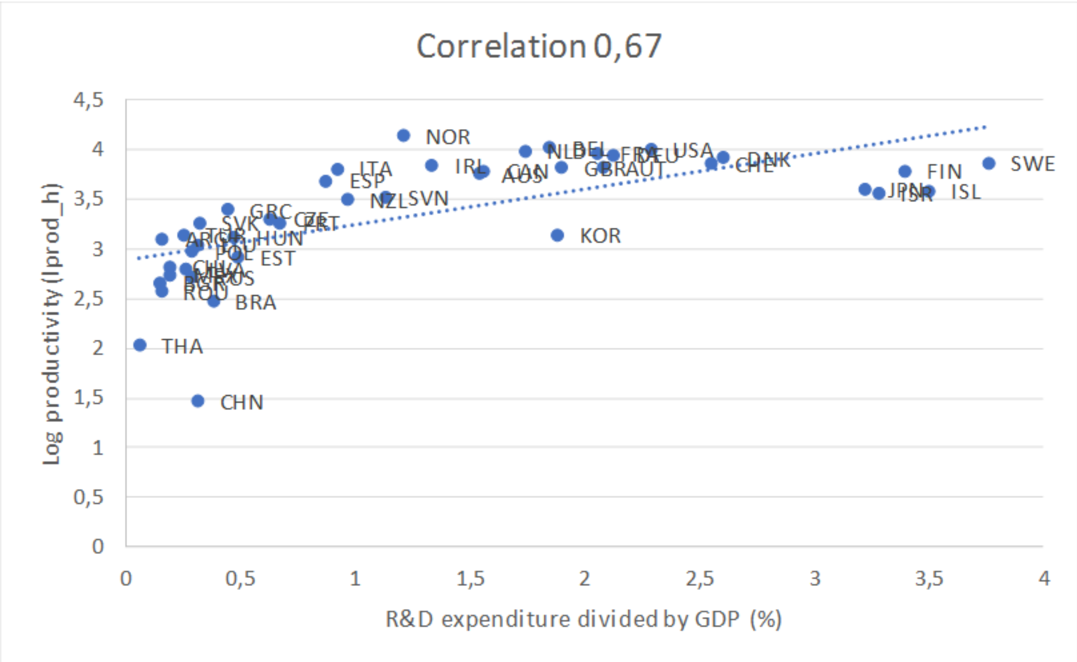
3. Data and variables

The time period of our sample is 1995–2011. This span is chosen to allow the inclusion of as many countries as possible. The chosen time interval allows us to use the OECD's Input-Output data and the PWT9.0 data set for a wide variety of countries. The availability of proper R&D data restricts the country set to 42 developed or advanced developing countries.

Our data originate from different sources. The most important sources are the Penn World Tables 9.0 (PWT9.0) by Feenstra, Inklaar and Timmer (2015) and the World Input-Output data (WIOD) by Timmer and Dietzenbacher, Los, Stehrer and de Vries (2015). Information is also obtained from OECD data sets, AMECO, World Development Indicators (WDI), Eurostat, Barro and Lee (2013), the Stockholm International Peace Research Institute (Sipri), World Mining Data by Reichl et al. (2016), the Energy Information Administration (EIA), the GeoDist data set by Mayer and Zignago (2011), the Center for World University Rankings (CWUR) and United Nations databases. The data source and construction of each variable is shown in Appendix 1, whereas Table 1 presents the list of variables, their abbreviations and short descriptions. Several missing observations are obtained using alternative data sources, and on rare occasions, we linearly interpolate missing values.

Next, we briefly discuss the character of the main variables. Our main interest is in labour productivity (prod), but for robustness, GDP per capita is regarded as an alternative. Based on modern growth theory, we let R&D intensity and enrolment ratios – that are constructed to correspond with the adult skill level (education) – impact economic growth. We discovered that the structure of the manufacturing industry – specifically, the share of consumer goods production (share of c-goods) – impacts productivity, either directly or through R&D intensity. These three variables - R&D intensity, education and the share of c-goods – are regarded as endogenous regressors whose time and cross-sectional variations are accounted for in the estimation of the model parameters.

Figure 1. R&D-intensity and labour productivity, time averages

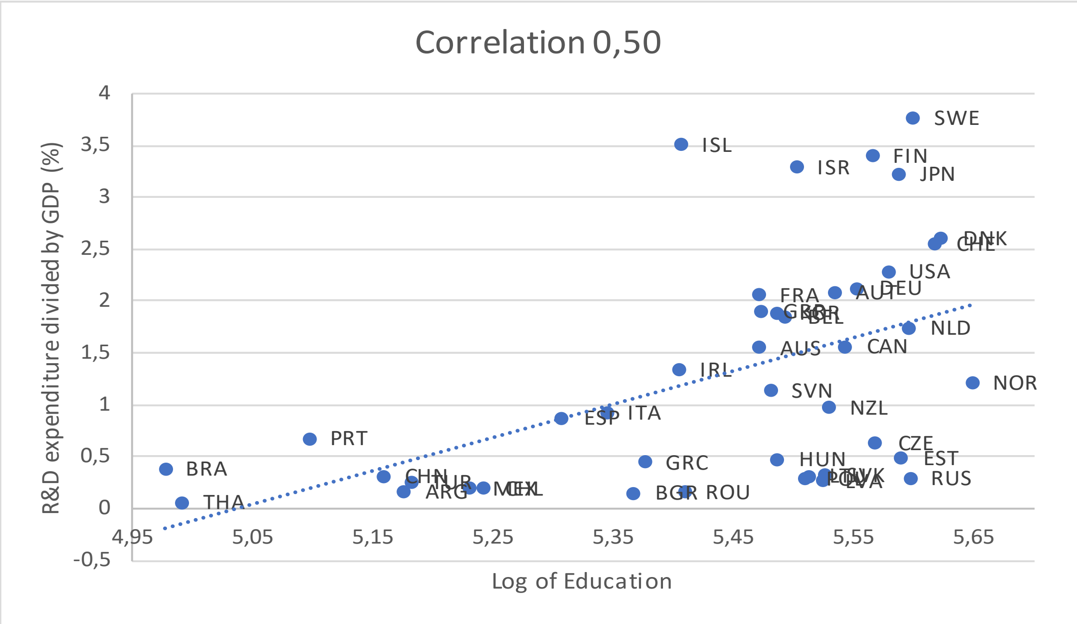


R&D intensity represents the effort devoted to research and development within a country. R&D intensity is measured in various ways in the literature, as noted by Madsen (2008). Here, we measure R&D intensity by dividing real R&D expenditure by real GDP. Therefore, nominal values of R&D expenditure are deflated to real values using the (2005=100) prices of gross fixed capital formation as a deflator, when this information is available. Otherwise, we use the GDP deflator. For certain countries, the availability of R&D expenditure data is rather limited; in these cases, linear interpolation is used to fill in the missing values (see Appendix 1). The correlation between R&D intensity and labour productivity is 0.67, which is quite strong, as expected.

To construct the education variable, we combine enrolment data with information from the OECD PIAAC study regarding adult skill levels.[11] The time variation of this variable reflects enrolment in higher education (secondary plus tertiary), data for which are obtained from Eurostat, OECD and World Bank data sets, and the cross-sectional variation resembles the differences among adult skill levels revealed by the OECD PIAAC study. The use of PIAAC results is motivated by the fact that enrolment variables do not account for differences in the quality of education among countries. Schooling systems and the requirements for higher degrees vary across countries, and this variation is not seen in data on formal levels of education. The construction of the education variable is described in detail in Appendix 1.

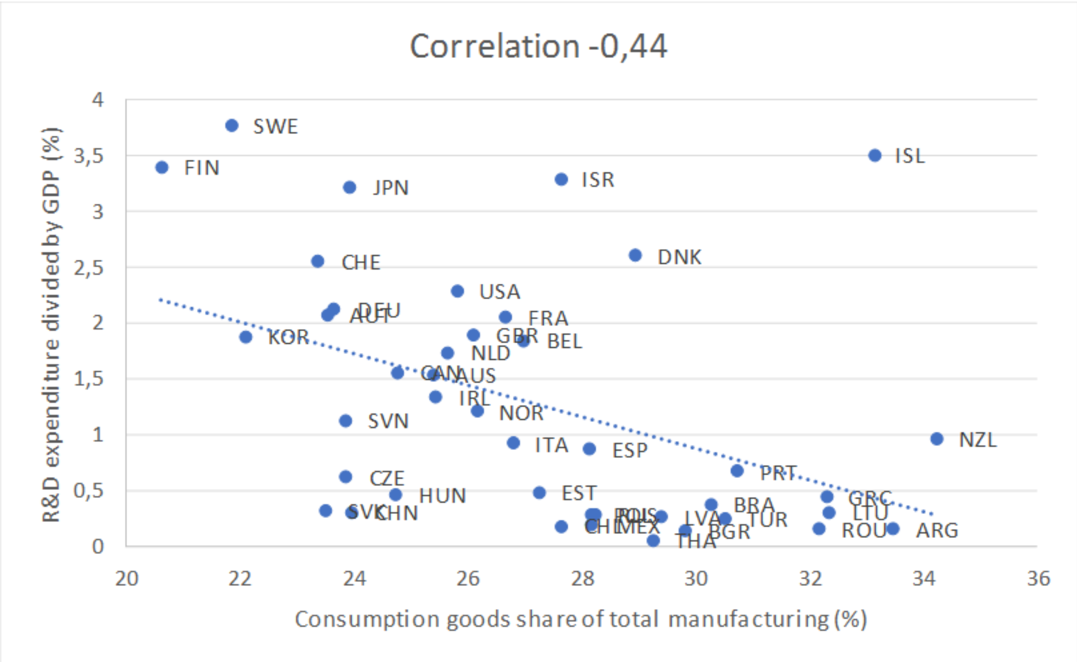
[11] In the tables of this study, this variable (which also reflects adult skills) is called "Education".

Figure 2. The level of education and R&D-intensity, time averages



The variable 'Share of c-goods' is a proxy for the structural composition of a country's manufacturing sector. Specifically, it represents the portion of a country's manufacturing sector that is devoted to producing goods that are used for final consumption, as opposed to products that are used as investments or to produce other goods (i.e., intermediate goods). We initially hypothesized that the technological requirements for the production of intermediate or investment goods are more demanding than those for the production of consumer goods. The correlation between productivity and the share of manufacturing devoted to consumption goods is -0.31. The correlation is not substantial, but it indicates that high productivity levels may be related to smaller shares of consumer goods production in total manufacturing. Moreover, the correlation between R&D intensity and the share of c-goods is negative and quite large, -0.44, which indicates that the greater the effort devoted by a country to research and development, the more that country produces intermediate and investment goods, which suggests that the production of these goods is more technologically demanding.

Figure 3. The industrial structure and R&D-intensity, time averages



The geographic variables used in our study are population size, internal density, and remoteness in relation to other countries. In the empirical analysis, we use time averages of each geographic variable to eliminate time variation. The restriction of our focus to cross-sectional variation is based on the need to ensure that the geographic variables are exogenous and thus independent of the economic progress being investigated.

For internal density within the country, we have constructed two alternative variables. The first is traditional density (Dens1), which is population divided by the square of the country's area. The second measure is the inverse for the average distance between persons distributed uniformly within the country.[12] This measure (Dens2) is obtained by dividing the population by the square of the country's area.

To obtain external density (centrality), we utilize distances provided by the GeoDist database of Mayer and Zignago (2011). In the GeoDist data set, the distance between different countries is calculated using the locations (based on latitudes and longitudes) between the most important cities (in terms of population) in any two countries. These distances are then weighted by appropriate mass variables, which in the external density variable is either population or GDP in the neighbouring countries.

The centrality variable Extdens1 is obtained by using population weights, whereas extdens2 uses GDP weights. We also find that (after all this calculation) Extdens2 is strongly negatively correlated with country area, which reflects the fact that the distance from a large country to its border area is typically long.

[12] According to Head and Maye (2002) the distance within the country is then a radius of a circle ($\sqrt{country's\ area/\pi}$) multiplied with a fixed scalar.

However, if economic interaction occurs near the borders, Extdens2 may undervalue the centrality of large countries. Therefore, we constructed a third measure, Extdens3, in which the impact of the country's area is removed. The measures used to assess the geographical centrality (remoteness) of a country are similar to those used in trade gravity models. One may wonder why we have three different variables for centrality. The answer is that there exists no unique genuine way to measure centrality, and consequently, the impact of centrality varies from one variable to another. Therefore, the use of a single centrality variable – for example, Extdens2 – would not yield an accurate picture of the robustness and persistency of the analysed impacts.

The natural resource variables in our data are as follows: percentage of land area covered by arable agricultural land (Agri); percentage of land area covered by forest (Forest); proven reserves of natural gas and oil divided by population (Gas&Oil); and total production of minerals divided by population (Minerals). A more detailed description of the construction of these variables is provided in Appendix 1 (see also Table 1). The endogeneity problem discussed above also affects the natural resource variables. For example, time variation in the Agri variable seems to be caused by urbanization, which is occurring in most countries. Economic progress typically reduces agricultural land area. However, due to variations in latitude and terrain, country differences in the time averages of the Agri variable are large and persistent. Due to the possible endogeneity in time variation, we consider only the time averages of these variables. To ensure the exogeneity of the geographic and natural resource variables, we use time averages from the time span 1990 – 1994 that precedes the estimation period 1995 – 2011.

The literature, which is referred to in the introduction, suggests that the impacts of R&D on productivity are conditional on the technological level of the country in question. It is then natural to hypothesize that geographic density and centrality, insofar as they grasp the impacts of agglomeration, impact productivity according to the country's stage of development. Owing to this, we interact geographic variables with the level of a country's labour productivity in 1990. For more discussion concerning the interaction terms, see section 4. Boschini et al. (2007 and 2013) discovered that the quality of institutions improves the impacts of natural resources. It is then natural to assume that these impacts are also contingent on the initial level of productivity. The estimated models includes interactive variables for natural resource variables, too. In the case where the interactive variables are not statistically significant, they are, however, omitted.

In certain models, we also control for the portion of military spending in GDP (Milit), which is classified as time variant and exogenous. University ranking (Univ) is also included in certain models as a time-invariant and exogenous variable.

Table 1. Data

Dependent variable	
Productivity (real (PPP) GDP divided by total hours) (log)	Prod
GDP per capita (real (PPP) GDP divided by population) (log) (alternative for prod)	GDP per capita
Time-variant endogenous variables	
Consumption goods share of total manufacturing (log)	Share of c-goods
R&D expenditure divided by real GDP (log)	R&D-intensity
Secondary plus tertiary enrolment rates adjusted with PIAAC levels (log)	Education
Military expenditure divided by GDP (log)	Milit
Time-invariant exogenous variables	
Population density (log)	Dens1
Geographical density (country's "own distance", population weigths) (log)	Dens2
Geographical centrality (population weights) (log)	Extdens1
Geographical centrality (GDP weigths) (log)	Extdens2
Geographical centrality (GDP weigths and area "corrected") (log)	Extdens3
Mean population (log)	Population
Percent of land area covered arable agricultural land (log)	Agri
Percent of land area covered by forest (log)	Forest
Proved reserves of natural gas and oil divided by population (log)	Gas&Oil
Total production of minerals divided by population (log)	Minerals
University rank index divided by population (log)	Univ
Cross-term for Population (variable multiplied by Prod90)	Pop*Prod90
Cross-term for Dens1 (variable multiplied by Prod90)	Dens1*Prod90
Cross-term for Dens2 (variable multiplied by Prod90)	Dens2*Prod90
Cross-term for Extdens1 (variable multiplied by Prod90)	Extdens1*Prod90
Cross-term for Extdens2 (variable multiplied by Prod90)	Extdens2*Prod90
Cross-term for Extdens3 (variable multiplied by Prod90)	Extdens3*Prod90
Cross-term for Agri (variable multiplied by Prod90)	Agri*Prod90
Cross-term for Forest (variable multiplied by Prod90)	Forest*Prod90
Cross-term for Gas&Oil (variable multiplied by Prod90)	Gas&Oil*Prod90
Cross-term for Minerals (variable multiplied by Prod90)	Minerals*Prod90
Deviation of cross-country mean productivity in year 1990	Prod90
Deviation of cross-country mean GDP per capita in year 1990 (alternative for Prod90)	Gdppop90

In Table 1, we provide a short description of the above variables. A more detailed description of their construction is given in Appendix 1. In Table A1 (in the Appendix), we report the summary statistics of our data and rank the countries in our sample based on different variables. The figures in Table A1 are log values of the averages over the years. In Table 3, we report the correlations between different variables.

4. Estimation

4.1. Estimation method

Our study considers the impacts of time-invariant geography and natural resources on industrial and input structure, R&D-intensity, adult skill levels, and labour productivity. The estimated models include both time-varying and time-invariant variables, which sets nonstandard requirements on the estimation. Consider the following model:

$$(1)\quad Y_{it} = X_{it}^{1}\beta^{1} + X_{it}^{2}\beta^{2} + Z_{i}^{1}\gamma^{1} + Z_{i}^{2}\gamma^{2} + \alpha_{i} + \varepsilon_{it},$$

where X_{it}^{1} is the vector of time varying exogenous variables, X_{it}^{2} is the vector of time varying endogenous variables, Z_{i}^{1} is the vector of time invariant exogenous variables, and Z_{i}^{2} is the vector of time invariant endogenous variables. In (1), α_i is the unobserved, panel-level random effect with zero mean and finite variance δ_{α}^{2}, and ε_{it} is the idiosyncratic error with zero mean and finite variance δ_{ε}^{2}. The error ε_{it} is uncorrelated with the columns of (X,Z,α). However, the latent individual effect α_i is correlated with the endogenous X_{it}^{2} and Z_{i}^{2}, due to which OLS and random effects (or GLS) yield biased and inconsistent estimates of the parameters β, γ, δ_{α}^{2} and δ_{ε}^{2}. The within estimator consistently estimates β^1 and β^2, but it removes time-invariant variables, meaning that neither γ^{1} nor γ^{2} can be estimated. Hausman and Taylor (1981) introduced a method to also obtain consistent and efficient estimates for γ^{1} and γ^{2} as long as some columns of (X,Z) are uncorrelated with α_i. It is also required that X_{it}^{1} be sufficiently correlated with Z_{i}^{2} and that the number of variables in X_{it}^{1} be at least as high as the number of variables in Z_{i}^{2}. The stages in the Hausman-Taylor estimation (HT) are as follows:

- Obtain the fixed effects estimates of $\hat{\beta}_{FE}^{1}$ and $\hat{\beta}_{FE}^{2}$ and a consistent estimate of residual variance σ_{ε}^{2}.
- Then, calculate within group-residuals $\hat{d}_{i} = \bar{Y}_{i} - \bar{X}_{i}^{1}\hat{\beta}_{FE}^{1} - \bar{X}_{i}^{2}\hat{\beta}_{FE}^{2} = Z_{i}^{1}\gamma_{1} + Z_{i}^{2}\gamma_{2} + \alpha_{i} + \bar{\varepsilon}_{i}$, where a superscript "-" refers to a group mean.
- Use $\bar{X}_{i}^{1}$ as instruments for Z_{i}^{2} to obtain 2SLS-estimates of γ^{1} and γ^{2} from the above equation, which explains the requirement k1 ≥ g2.

To derive consistent estimates $\hat{\beta}_{FE}^{1}$ and $\hat{\beta}_{FE}^{2}$, all the between variation is neglected, which makes these estimates inefficient. This neglect is also reflected in $\hat{\gamma}_{2SLS}^{1}$ and $\hat{\gamma}_{2SLS}^{2}$, making them inefficient, as well. Accordingly, the initial estimates so far obtained for β^1, β^2, γ^{1} and γ^{2} are used to construct weighted instrumental variable estimators for these parameters. This process involves the following steps:

- The residual variance (σ_d^2) in the previous step (when the equation for $\hat{d}_i$ was estimated) is $\sigma_d^2 = \frac{\sigma_\varepsilon^2}{T} + \sigma_\alpha^2$, from which it is obtained that $\sigma_\alpha^2 = \sigma_d^2 - \frac{\sigma_\varepsilon^2}{T}$. On the right hand side of this equation, σ_ε^2 was obtained in the fixed effects estimation.
- The weight for a feasible GLS-estimator is then $\hat{\theta} = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}$.
- Using $\hat{\theta}$, we obtain GLS transforms $w_{it}^* = w_{it} - \hat{\theta}\bar{w}_{it}$, where $w_{it} = (X_{it}^1, X_{it}^2, Z_i^1, Z_i^2, \alpha_i, \varepsilon_{it})$.
- Equation (1) is then estimated in the form $Y_{it}^* = X_{it}^{1*}\beta^1 + X_{it}^{2*}\beta^2 + Z_i^{1*}\gamma^1 + Z_i^{2*}\gamma^2 + \alpha_i^* + \varepsilon_{it}^*$, using instruments $X_{it}^k - \bar{X}_i^k$, $\bar{X}_i^k$ and Z_i^1 for endogenous X_{it}^{2*} and Z_i^{2*}, where k=1,2.

Ameniya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989) propose additional instruments such as all time periods of X_{it1} (in the last step above) to improve the efficiency of the HT estimator.

However, in the models, we estimate all time-invariant geography variables, and time-invariant variables that describe the possession of certain types of natural resources are considered exogenous. Furthermore, since in our setting all time-invariant variables are considered as exogeneous and all time-varying variables as endogenous, we are estimating a special case of the HT estimator in which both X_1 and Z_2 are "empty", and HT and AM produce the same standard errors.[13] As Breusch et al. (2011b) note, HT under the above exogeneity assumptions has a simple IV representation where the deviations from group means of the endogenous time varying variables serve as instruments for the endogenous time-varying variables, and the exogenous time-invariant variables serve as their own "instruments". Therefore, it is simple to implement, for example, the Arellano (1987) type clustered robust covariance matrix of parameters that accounts for heteroskedasticity and serial correlation.

We apply the Hausman (1978) test regarding the suitability of either a fixed effects panel estimator or a random effects estimator for each equation we estimate. According to these test results, in almost every specification, the estimation should be executed with fixed effects estimation because the time-varying explanatory variables are endogenous. Although we discovered that for a few specifications where R&D-intensity is the dependent variable, the tests do not support fixed effects estimation, we estimate all our models as if the time-varying regressors were endogenous. Intuitively, it would be rather inconsistent to define any of the potentially endogenous variables (i.e., R&D intensity, c-goods or education) as exogenous. Therefore, we choose to utilize HT-estimation over random effects estimation.

[13] Under these exogeneity assumptions, the HT estimator produces the same coefficients as the so-called "fixed effects vector decomposition (FEVD)" estimator (Plümper and Troeger, 2007) as pointed out by Breusch et al. (2011a). The important difference between HT and FEVD is that Plümper and Troeger (2007) draw their standard errors from an incorrect covariance matrix (Greene, 2011; Breusch et al., 2011b). Breusch et al. (2011a) state that there is no reason why the HT estimator should not be used given the above exogeneity assumptions since the estimator is still well defined.

Another issue in the estimation of cross-country panel regressions is whether to add time fixed effects into the regression. The study of Serlenga and Shin (2007) regarding bilateral trade flows among 15 European countries is somewhat similar to our study. Their model also includes time-invariant variables. They discuss whether one should include time fixed effects in the model. They find that the estimates that include (homogeneous) time fixed effects are unreliable compared to the standard Hausman and Taylor (1981) and CCEP-HT[14] estimates. Accordingly, we do not include time fixed effects in our estimation.

There is yet one issue that requires further attention. The inclusion of interaction terms results in heavy multicollinearity, especially since we add the direct effect of Prod90. We add Prod90 as a regressor because Bernhardt and Jung (1979) note that in a polynomial regression, inferior order terms should be included to achieve meaningful results that are not affected by the chosen scale of the variables. We handle the problem of multicollinearity as follows. We transform all the time-invariant variables into the form $\log(\frac{z_i}{\bar{z}_i})$ except for Prod90, and we have the following form: $prod90_i - \overline{prod90}_i$.[15] Furthermore, all the interactive terms are then in the form $\log(\frac{z_i}{\bar{z}_i}) * (prod90_i - \overline{prod90}_i)$, where the upper bar denotes the average of the variable. Prod90 is not in log form since this reduces the multicollinearity even more. We treat $(prod90_i - \overline{prod90}_i)$ as exogenous. This is motivated by the fact that 1990 lies 5 years outside our sample.

The above transformations greatly reduce multicollinearity. However, according to the VIF (variance inflating factor) analysis, there is still some multicollinearity left when considering the full model with all interaction terms.[16] To achieve less inflated standard errors, we remove those interaction terms that are clearly not significant. We do this by excluding terms one by one, starting from the most insignificant, until we are left with a specification where the remaining interaction terms are at least approximately near the 0.2 significance level.

4.2. The estimated model

The estimated model reflects the framework of modern growth theory. R&D investments and education are thus seen as drivers of growth in developed countries. The more traditional production factors – fixed capital and labour input - are omitted. Instead, we add a variable that describes the share of consumer goods production in total manufacturing. Although the share of manufacturing in GDP exhibits a declining trend,

[14] CCEP-HT is an estimation method developed by Serlenga and Shin (2007). This estimation method combines the correlated common effect pooled (CCEP) estimation approach advanced by Pesaran (2006) and the Hausman and Taylor (1981) estimation approach.

[15] This transformation has not been done to Extdens3 since its construction deviates around zero, see Appendix 1.

[16] The VIF is mostly under 10; however, in some full models, this value is little bit above 10 for some variables (often for Prod90). After we exclude the most insignificant variables, the VIF values for all the variables are under 10.

this variable seems to be a good indicator of the technological stage of the country considered. In more advanced countries with high R&D intensity, this share is below average.

In analysing the impacts of geography and natural resources, we have estimated linear equations for labour productivity (Prod), R&D intensity (Rd), education level (Ed) and the share of consumer goods production in manufacturing (Cg).[17] These equations are estimated separately, and we first report the results using the results of Hausman-Taylor estimator. [18] We estimate the following linear equations:

(e1) $Prod\ = a0 + a1 * Rd + a2 * Cg + a3 * Ed + a4 * Geo + a5 * Geo * prod90 + a6 * Nat + \varepsilon_1$

(e2) $Rd = b0 + b1 * Cg + b2 * Ed + b3 * Geo + b4 * Geo * Prod90 + b5 * Nat + \varepsilon_2$

(e3 $Cg\ = c0 + c1 * Rd + c2 * Ed + c3 * Geo + c4 * Geo * prod90 + c5 * Nat + \varepsilon_3$

(e4) $Ed\ = d0 + d1 * Rd + d2 * Cg + d3 * Geo + d4 * Geo * Prod90\ + d5 * Nat + \varepsilon_3$.

where

Geo = geographic variable (Dens, Extdens and Pop)

Nat = natural resources (Agri, Forest, Gas&Oil and Minerals)

ε_i = error term.

In equations (e1) – (e4), the variables on the left-hand side are defined as endogenous. Geographic and natural resources variables are considered exogenous. The geographic and natural resource variables thus affect the endogenous variable directly (when the other endogenous variables are regarded as given) and indirectly via the other endogenous variables. We first report the estimation results when each equation is estimated separately. After that, we calculate the total impacts of geographic and natural resource variables considering equations (e1) – (e4) as a simultaneous equation system. In that case, we take into account the impacts of the exogenous variables that move through other endogenous variables.

The definitions of variables as either endogenous or exogenous originate more from economic theory than from statistical testing. If endogeneity is assumed, it follows that the random effects model - which implicitly assumes that all the regressors are exogenous - is excluded. The fact that country-specific fixed effects would remove all impacts of time-invariant variables also makes this estimator useless for the purposes of this study. Based on the aforementioned factors, the HT-estimator is deemed appropriate for this study.

In analysing the impacts of natural resources and geography on productivity and other endogenous variables, we try to emphasize the impacts of permanent and long-lasting features of the environment. The use of

---

[17] These variables are mostly in log form. They are described in more detail in Table 1 and in Appendix 1.

[18] Note that the reported coefficients are the same as from standard HT estimation, but SE's are the type of Arrellano (1987); see section 4.

Hausman-Taylor estimator sets fixed effects estimates for time-varying variables in the first stage and fixes the coefficients of time-invariant variables in the subsequent stages. Therefore, this estimator is appropriate for the model that also includes natural resource and geographic variables. Some of these variables are originally time invariant, whereas others vary over time. Certain variables – such as population and its density – even share the same time variation. The time variation of these variables during the studied time period may correlate with certain endogenous variables, such as productivity and R&D intensity. In certain cases, this variation appears to be an outcome of economic progress rather than its cause. For example, rapid economic growth in a country tends to increase population size and density in that country.[19] Additionally, the agricultural land area in many countries has decreased due to urbanization and population growth. Therefore, the time variation of these variables is not necessarily exogenous with respect to the endogenous variables considered. In contrast, country-specific differences between the time averages of these variables are long-lasting and persistent. For example, the relative difference in population sizes between any two countries in the data has been approximately the same for decades or even centuries. The same is true for the density and centrality variables. Therefore, we consider only the time averages of natural resource and geographic variables. In so doing, we ensure that only persistent differences are taken into account, and we also avoid the hazards of the chosen econometric procedure. Restricting variation to cross-sectional variation would yield substantially different assessments of the impacts of this variable. Giving so much weight to cross-sectional variation in the data may, however, lead to some other pitfalls, including the omitted variable bias. In the model considered, the coefficients of explanatory endogenous variables are sensitive to the variables selected.

In the econometric approach chosen, we let the endogenous variables that describe R&D intensity, education level, productivity and production structure vary in time. Either labour productivity or GDP per capita only is used as a dependent variable. Letting R&D intensity, education level (adult skills) and the production structure explain productivity or GDP is in line with modern growth theory. Because we are interested in the impacts of geography and natural resources, we also consider how the other endogenous variables are affected by the remaining endogenous regressors and by all the exogenous variables. For example, R&D intensity is explained by education, production structure, and all the exogenous variables.

In the models, the reported impacts of the natural resources and geographic variables are conditioned by the labour productivity in 1990 (Prod90). Insofar as the interactive variable, which is the variable in question multiplied by Prod90, is statistically significant, the direct effect is actually the net effect of the impact of the

---

[19] In empirical growth studies that use country data, the endogeneity of apparent causes of growth is a major problem. This problem is also characteristic of the models in which growth is explained by institutions (see, for example, Glaeser et al.,2004).

variable in question plus the impact of the interactive term. Table A1 reports the value of Prod90 and the other variables for each country.

5. Results

5.1 The direct impacts of natural resources and geography

R&D intensity

In all models, the education variable positively impacts R&D intensity and negatively impacts the share of consumer goods production. The effects of geographic variables on R&D reflect market reactions and possible policy interventions. Among geographic variables, internal density (Dens1 and Dens2) has a positive effect on R&D intensity, indicating that the positive externalities of agglomeration occur. Only in model 1 (the model in the first column in Tables 2 – 5), the interaction term (Dens1*Prod90) is statistically significant. In that model, the direct net impact (of Dens1 and Dens1*Prod90) is positive for countries whose Prod90 is below the 88^{th} percentile.

The direct impact of external density or centrality (Extdens1 and Extdens2) on R&D intensity is negative in all the models. Negativity hints that a tendency exists that works against the adverse impacts of remoteness by promoting R&D investments through markets or public measures. How is this explained? One possible explanation for this tendency is that remote countries do not attract knowledge in the form of foreign investments, which forces them to invest in R&D themselves. It is also possible that geographic closeness to neighbouring countries allows a country's manufacturers to subcontract with principal firms in the neighbouring country. In this model, a country need not make R&D investments as large as those required when the country must rely on self-making in the most stages of production. It is thus logical that the latter pattern could lead to higher R&D intensity. Again, only in model 1 does the interaction term (extdens3*prod90) have a statistically significant explanation, but the net impact is still negative for almost all the countries.

Population size (Pop) has a negative impact on R&D intensity in models 1 and 4 (the models in the first and fourth columns in Tables 2 – 5). In model 2, the net impact is negative when Prod90 is above the median level. In model 3, negativity requires that Prod90 is above the 70^{th} percentile. That population's impact is negative especially when the initial productivity (Prod90) is high indicates that there exists a tendency to compensate real or imaginary defects of smallness by encouraging R&D investments. The basic reasons for this may be similar to those for centrality.

Table 2, dependent variable R&D-intensity

	(1)	(2)	(3)	(4)
Time-varying endogenous variables				
Share of c-goods	-0.756*	-0.756*	-0.756*	-0.756*
	(0.354)	(0.354)	(0.354)	(0.354)
Education	0.693**	0.693**	0.693**	0.693**
	(0.224)	(0.224)	(0.224)	(0.224)
Time-invariant exogenous variables				
Agri	-0.289´	-0.217´	-0.212	-0.212
	(0.157)	(0.128)	(0.131)	(0.131)
Gas&Oil	-0.059**	-0.063**	-0.064**	-0.064**
	(0.017)	(0.017)	(0.017)	(0.017)
Forest	-0.145*	-0.086	-0.089	-0.088
	(0.058)	(0.064)	(0.065)	(0.064)
Minerals	0.056	0.091*	0.094*	0.094*
	(0.040)	(0.040)	(0.039)	(0.039)
Population	-0.032	0.006	0.063	-0.117
	(0.055)	(0.046)	(0.056)	(0.086)
Dens1	0.434**	0.230**	0.180*	-
	(0.086)	(0.084)	(0.075)	
Dens2	-	-	-	0.360*
				(0.144)
Extdens1	-0.758**	-	-	-
	(0.257)			
Extdens2	-	-0.267´	-	-
		(0.155)		
Extdens3	-	-	-0.272´	-0.264´
			(0.152)	(0.151)
Agri*Prod90	0.035**	**0.024**	0.022	0.023
	(0.013)	**(0.015)**	(0.015)	(0.015)
Minerals*Prod90	0.011*	0.009*	0.009*	0.009*
	(0.004)	(0.004)	(0.004)	(0.004)
Pop*Prod90	-	-0.010*	-0.009*	-0.010*
		(0.004)	(0.004)	(0.004)
Dens1*Prod90	-0.031**	-	-	-
	(0.009)			
Extdens1*Prod90	0.070**	-	-	-
	(0.026)			
Prod90	0.093**	0.091**	0.091**	0.091**
	(0.015)	(0.009)	(0.009)	(0.009)
N	714	714	714	714

Arrellano (1987) type robust standard errors in parentheses.

Significance levels: . <0.1 * <0.05 **<0.01.

Bolded values indicate that coefficients are significant at least at the 0.1 level with basic HT standards errors.

Of the natural resources, gas and oil have a strong negative impact on the R&D intensity. However, Forest and Agri variables also have a negative effect. For few countries with the highest Prod90 does the impact of Agri turn positive. It is reasonable to regard these impacts as due exclusively to market factors rather than as a combination of market factors and policy actions. In essence, this impact is similar to the so-called "natural resource curse" discovered by Sachs and Warner (1995 and 2001). In contrast to other natural resources, minerals, however, seem to raise R&D intensity. That this impact strengthens with Prod90 is in line with the findings of Boschini et al. (2007 and 2013).

The share of consumer goods production

The share of consumer goods production (Share of c-goods) is negatively related to both R&D intensity and education level. This result indicates that the quality requirements for consumer goods are lower than those for other types of goods, which are purchased by firms and not by consumers. Attributes other than quality – for example, image, fashion and local taste – may have a remarkable impact on the demand for consumer goods.

The home market effect (HME) of economic geography would suggest that large population and dense markets favour the production of intermediate and investment goods, whose production is typical of large fixed costs and high trade costs.[20] According to another view, the large population and dense markets could instead favour the production of consumer goods if consumers preferred local goods and if the consumer goods production exhibited increasing returns to scale. The results show that in models 1 and 2, population size increases the share of intermediate and investment goods production, which for this part supports the HME hypothesis. [21] In models 3 and 4, it is internal density (Dens1 and Dens2) that pushes the share of intermediate and investment goods production up. This result is also in line with the HME hypothesis.

The external density in our model extends the local demand of a good on grounds of the distance from the own country in a similar way as in Davis and Weinstein (2003) and Behrens et al. (2004). That external density (Extdens1, Extdens2 and Extdens3 plus the respective interactive variables) also favours the production of investment and intermediate goods brings some additional proof for the existence of the HME. Therefore, we can say that the production of intermediate and investment goods tends to be located in the heavily populated and centrally located countries.

[20] This was the original hypothesis of Krugman (1980), for which Davis and Weinstein (2003), Hanson and Xiang (2004) and Crozet and Triofetti (2008), for example, found empirical support.

[21] That no impact is discovered in models 1 and 4 may partly be explained by the fact that the data in our study are much more aggregated than data on industrial structure that are customarily analysed to prove the possible existence of HME. Furthermore, our variable of country size is the population and not GDP, which is used in the aforementioned studies in the field of international trade.

Table 3, dependent variable the Share of c-goods

	(1)	(2)	(3)	(4)
Time-varying endogenous variables				
R&D-intensity	-0.096**	-0.096**	-0.096**	-0.096**
	(0.036)	(0.036)	(0.036)	(0.036)
Education	-0.420**	-0.420**	-0.420**	-0.420**
	(0.074)	(0.074)	(0.074)	(0.074)
Time-invariant exogenous variables				
Agri	-0.031	0.002	0.005	0.004
	(0.026)	(0.033)	(0.034)	(0.034)
Gas&Oil	0.004	0.004	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Forest	-0.028´	-0.022	-0.022	-0.022
	(0.015)	(0.017)	(0.018)	(0.018)
Minerals	-0.012	-0.025**	-0.021*	-0.021*
	(0.009)	(0.009)	(0.008)	(0.008)
Population	-0.027*	-0.024*	-0.008	0.018
	(0.012)	(0.011)	(0.011)	(0.018)
Dens1	0.004	-0.015	-0.026´	-
	(0.021)	(0.015)	(0.015)	
Dens2	-	-	-	-0.052´
				(0.030)
Extdens1	**-0.105**	-	-	-
	(0.070)			
Extdens2	-	**-0.059**	-	-
		(0.036)		
Extdens3	-	-	**-0.054**	**-0.055**
			(0.037)	**(0.037)**
Agri*Prod90		-0.003	-0.004	-0.004
		(0.003)	(0.003)	(0.003)
Forest*Prod90	-0.003´	-0.006**	-0.006**	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)
Minerals*Prod90	-	-0.002´	-0.001	-0.001
		(0.001)	(0.001)	(0.001)
Pop*Prod90	-	-	0.001	0.001
			(0.001)	(0.001)
Dens1*Prod90	0.002	-	-	-
	(0.001)			
Extdens1*Prod90	-0.010*	-	-	-
	(0.004)			
Extdens2*Prod90	-	-0.006*	-	-
		(0.002)		
Extdens3*Prod90	-	-	-0.006´	-0.006´
			(0.003)	(0.003)
Prod90	0.007**	0.002	**0.006**	**0.006**
	(0.002)	(0.004)	**(0.004)**	**(0.004)**
N	714	714	714	714

Arrellano (1987) type robust standard errors in parentheses.
Significance levels: . <0.1 * <0.05 **<0.01.
Bolded values indicate that coefficients are significant at least at the 0.1 level with basic HT standards errors.

The impact of natural resources on industry structure is clear. Wood and minerals are often processed to become intermediate products on sites that are located close to forestland and mines. Typically, some manufacturers of machinery and equipment used to process these raw materials have been established on the same sites. These factors explain why the mine and forest variables decrease the share of consumer goods production. In contrast, agricultural land area and gas and oil reserves have hardly any impact on the production structure.

Education level

Education level is positively impacted by R&D intensity, whereas the impact of Share of c-goods is negative. In all models and for nearly all countries, internal density (Dens 1 and Dens2 plus the respective interactive variables) decreases education level. In Appendix A2.3, we hypothesized that these relations may be explained by the immigration habits of highly skilled persons. The high net immigration of talents, typical of densely populated and highly developed countries, may decrease the social returns of schooling and thus result in lower investments in education. It is also possible that authorities tend to compensate for the impact of low density and the related low level of agglomeration by investing more in education than is socially optimal in the short term. With this strategy, a stubborn attitude towards education is later rewarded by a growing number of highly skilled people. However, this kind of behaviour is not easily rationalized. One might also question why all countries do not act in this way.

According to the results reported in Table 4, the variables for external density – Extdens1 and Extdens2 – have a positive impact on education when the initial productivity is low. However, for the countries whose Prod90 is above the 55^{th} – 60^{th} percentile, this impact turns negative. In poorer countries, households and the public sector may have no resources to compensate for the disadvantage caused by remote location. Large population seems to deteriorate education. In model 4, though, this impact is of opposite sign for the wealthiest countries.

Among natural resource variables, Gas&Oil has a positive effect on education level. Forest has no impact on education level, and Agri at the highest level has a slight negative impact. The strong positive impact of gas and oil reserves may be an indication that a portion of the private and public income from gas and oil sales is channelled into schooling. The clear negative impact of Minerals may be the result of low educational requirements in the industries that mine and refine minerals.

Table 4, dependent variable Education

	(1)	(2)	(3)	(4)
Time-varying endogenous variables				
R&D-intensity	0.134**	0.134**	0.134**	0.134**
	(0.047)	(0.047)	(0.047)	(0.047)
Share of c-goods	-0.635**	-0.635**	-0.635**	-0.635**
	(0.095)	(0.095)	(0.095)	(0.095)
Time-invariant exogenous variables				
Agri	-0.003	-0.027	-0.038	-0.039
	(0.030)	(0.026)	(0.026)	(0.026)
Gas&Oil	0.020**	0.024**	0.023**	0.023**
	(0.004)	(0.005)	(0.005)	(0.005)
Forest	0.011	0.006	0.005	0.005
	(0.014)	(0.017)	(0.019)	(0.019)
Minerals	-0.033**	-0.050**	-0.045**	-0.046**
	(0.009)	(0.011)	(0.011)	(0.010)
Population	-0.036**	-0.043**	-0.051**	-0.006
	(0.012)	(0.014)	(0.013)	(0.022)
Dens1	-0.095**	-0.056**	-0.044*	-
	(0.027)	(0.022)	(0.020)	
Dens2	-	-	-	-0.090*
				(0.038)
Extdens1	0.117´	-	-	-
	(0.060)			
Extdens2	-	0.063´	-	-
		(0.034)		
Extdens3	-	-	0.079*	0.078*
			(0.036)	(0.036)
Agri*Prod90	-0.007*	-	-	-
	(0.003)			
Minerals*Prod90	-0.003**	-0.002*	-0.002*	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Pop*Prod90	-	-	0.002**	0.002**
			(0.001)	(0.001)
Dens1*Prod90	0.009**	-	-	-
	(0.002)			
Extdens1*Prod90	-0.032**	-	-	-
	(0.007)			
Extdens2*Prod90	-	-0.009**	-	-
		(0.003)		
Extdens3*Prod90	-	-	-0.011**	-0.011**
			(0.003)	(0.003)
Prod90	-0.012*	-0.013*	**-0.008**	**-0.008**
	(0.006)	(0.006)	**(0.005)**	**(0.005)**
N	714	714	714	714

Arrellano (1987) type robust standard errors in parentheses.

Significance levels: . <0.1 * <0.05 **<0.01.

Bolded values indicate that coefficients are significant at least at the 0.1 level with basic HT standards errors.

Labour productivity

Regarding the endogenous variables, R&D intensity has a positive impact on labour productivity; the share of consumer goods production has a negative impact on labour productivity; and education level has no direct non-zero impact on labour productivity in the models presented in Table 5.

Table 5 shows that internal and external density and population size most often have a negative direct impact on productivity. Only for the richest countries do the direct net impacts of internal density become close to zero or slightly positive. This and the fact that the negative direct net impact of population size becomes diluted and may even turn positive when Prod90 increases indicate that dense and large countries with high enough Prod90 tend to benefit from cost savings in transportation and infrastructure.

Of the natural resources, the agricultural land area per inhabitant typically has a positive direct effect on productivity, although if Prod90 climbs above the 70^{th} – 75^{th} percentile, this impact turns negative. Forest and minerals affect productivity primarily negatively in the model framework of Table 5. Oil and gas reserves have no direct productivity impact.

5.2. The total impacts of geographic and natural resource variables

In the calculations described below, the total impact of a particular geographic or natural resource variable is the sum of the impact of this variable and its interactive term and the impacts that materialize through the other endogenous variables in models (1) – (7). The former impact is called a direct impact, whereas the latter is called an indirect impact. The total impact of geography and natural resources, being the sum of direct and indirect impacts, is calculated by solving the estimated equation system (e1) – (e4) for each endogenous variable (R&D intensity, Share of c-goods, Education and Prod), although we have not estimated these equations as a simultaneous equation system. In Table 6, we report the calculated total impacts of the geographic and natural resource variables on the original values of the endogenous variables when a geographic or natural resource variable changes from a median level to the 5^{th} highest level (approximately the 88^{th} percentile). In these calculations, the impacted endogenous variables are assumed to be at the median level. The calculations in Table 6 are derived from the model located in the fourth columns (model 4) of Tables 2 – 5 and, in parentheses, the model that is located in the third columns (model 3) of Tables 2 – 5.

Table 5, dependent variable Prod

	(1)	(2)	(3)	(4)
Time-varying endogenous variables				
R&D-intensity	0.151*	0.151*	0.151*	0.151*
	(0.067)	(0.067)	(0.067)	(0.067)
Share of c-goods	-1.445**	-1.445**	-1.445**	-1.445**
	(0.188)	(0.188)	(0.188)	(0.188)
Education	0.081	0.081	0.081	0.081
	(0.120)	(0.120)	(0.120)	(0.120)
Time-invariant exogenous variables				
Agri	0.093	0.253**	0.260**	0.278**
	(0.080)	(0.077)	(0.077)	(0.069)
Gas&Oil	0.015	0.001	-0.001	-0.004
	(0.011)	(0.012)	(0.012)	(0.011)
Forest	-0.060´	-0.059*	-0.060*	-0.051´
	(0.035)	(0.028)	(0.028)	(0.030)
Minerals	-0.001	0.007	0.018	0.028
	(0.021)	(0.018)	(0.017)	(0.021)
Population	-0.101**	-0.104**	**-0.058**	0.001
	(0.033)	(0.034)	**(0.038)**	(0.057)
Dens1	0.018	-0.015	-0.051	-
	(0.048)	(0.033)	(0.037)	
Dens2	-	-	-	-0.106
				(0.077)
Extdens1	-0.196	-	-	-
	(0.139)			
Extdens2	-	-0.197**	-	-
		(0.057)		
Extdens3	-	-	-0.174**	-0.169**
			(0.061)	(0.060)
Agri*Prod90	-0.014*	-0.028**	-0.030**	-0.031**
	(0.007)	(0.007)	(0.008)	(0.008)
Gas&Oil*Prod90	-	0.002*	0.002*	0.002*
		(0.001)	(0.001)	(0.001)
Minerals*Prod90	-0.005´	-0.006**	-0.005*	-0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
Pop*Prod90	0.011**	0.013**	0.010**	-
	(0.004)	(0.004)	(0.004)	
Dens1*Prod90	-	-	**0.006**	-
			(0.004)	
Dens2*Prod90	-	-	-	0.019**
				(0.006)
Extdens2*Prod90	-	0.015*	-	-
		(0.006)		
Extdens3*Prod90	-	-	0.014*	0.013*
			(0.006)	(0.006)
Prod90	0.024**	0.036**	0.034**	0.030**
	(0.009)	(0.008)	(0.009)	(0.007)
N	714	714	714	714

Arrellano (1987) type robust standard errors in parentheses.
Significance levels: . <0.1 * <0.05 **<0.01.
Bolded values indicate that coefficients are significant at least at the 0.1 level with basic HT standards errors.

Internal density

The total impact of internal density (Dens1 and Dens2 in models 3 and 4) on R&D intensity is strong and positive in all models and for almost all countries. Only in model 1 is the concerned impact negative for the five countries with the highest Prod90. The positive total impact is a result of a positive direct impact on R&D intensity and a negative direct impact on the share of consumer goods production. Internal density has a positive total effect on education level in models 3 and 4. This is explained by the indirect positive impact via R&D intensity. In model 2, this impact is close to zero, and in model 1, it is mostly negative. In the latter models, the negative direct impact in the equation (4) seems also to explain this result. Internal density (Dens2 and Dens1) also tends to lower the share of consumer goods production in models other than model 1. In the later models, this result holds for countries whose Prod90 is below the 70th percentile. Finally, internal density (Dens2) positively affects productivity.[22] This outcome holds for all countries in model 2 and for countries with Prod90 above the 30th percentile in model 5 and above the 20th percentile in model 3. In model 1, the total impact on productivity turns negative when Prod90 rises above the 80th percentile. Positivity is a result of the direct positive impact on productivity and R&D intensity and of the tendency of Dens2 and Dens1 to lower the share of consumer goods production. The calculations reported in Table 6 show that in absolute terms, the percentage impact of internal density on the original values of the endogenous variables are remarkable in models 3 and 4.

External density

External density (Extdens3) has mostly a negative total impact on R&D intensity. In models 3 and 4, this does not hold for the countries with very low Prod90 and in model 1 the countries whose Prod90 is above the 76th percentile. Negativity is a consequence of the negative direct impact on R&D intensity and of its negative impact via education level. The counter-impact, which occurs when external density lowers the share of consumer goods production, is not sufficiently strong.

The total effect of external density also lowers the share of consumer goods production (except in model 1 in a few countries with the lowest Prod90). This result is consistent with the corresponding direct impact. The total impact of centrality (variables Extdens3 or Extdens2) on education is positive for countries with Prod90 below the 75th percentile, which is also in line with the direct impact shown in the education equation. The total impact of centrality on productivity is negative when Prod90 is at the median level. In most models (model 1, 3, and 4), this impact turns positive when Prod90 rises above the 55th - 62nd percentile. In model 2,

[22] This result is in line with the findings of Gallup et al. (1999), although their data also included numerous developing and underdeveloped countries. Gallup et al. (1999) separately considered coastal density and inland density. The former had a strong positive impact on GDP growth in the period 1965-1990. Sachs and Warner (1995) also found that population density (the inverse of their land variable) positively affects GDP growth (between 1979 and 1989).

though, the concerning impact turns positive when Prod90 drops below the 45th percentile. The direct impacts via other endogenous variables on productivity go in opposite directions. Negativity is supported by the direct impact of external density on R&D intensity.

Population size

In models 1 and 4, population size in its entirety negatively affects R&D intensity. In models 2 and 4, this is valid only when Prod90 is above the median level. Additionally, population mostly tends to increase the share of consumer goods production and to lower the level of adult skills. These impacts explain why the total impact of population size lowers productivity in all models. However, in model 1, this does not apply in countries with very high values of Prod90. All in all, this is in contradiction with the HME hypothesis, according to which a large population supports industries other than mass production and therefore also productivity in comparison with countries with small population.

The calculations presented in Table 6 show that the results related to density and population are, however, very sensitive to variable choices. Replacing Dens2 with Dens1 increases the impact of population size on R&D intensity and, respectively, decreases the impact on the share of consumer goods production. In net, this kind of shift of variables would also decrease the total impact on labour productivity. We can also see from Table 6 that in model 4 (as compared with model 3), the strong positive impact of density is counterbalanced by the remarkable negative impact of population size (see Table 6).

Table 6. Total impacts (in percent) when endogenous variables and Prod90 are at the median level and the geographic or natural resource variables change from the median level to the 5th highest level in the model in the 4th column (in the model of the 3rd column)

	R&D-intensity	Share of c-goods	Education	Prod
Pop	-42.7 (-1.4)	19.8 (-0.6)	-17.4 (-10.1)	-30.2 (-17.3)
Dens2 (Dens1)	81.9 (31.7)	-13.8 (-1.8)	6.9 (3.5)	20.9 (10.3)
Extdens3	-5.8 (-6.4)	-5.5 (-5.4)	7.3 (7.1)	-1.5 (-2.1)
Agri	-10.3 (-10.1)	2.9 (3.1)	-4.3 (-4.4)	1.4 (0.5)
Forest	-4.5 (-4.6)	-1.7 (-1.6)	0.8 (0.8)	-2.0 (-2.7)
Minerals	30.8 (30.7)	-6.1 (-6.2)	-4.1 (-4.0)	21.5 (18.7)
Gas	-14.7 (-14.6)	0.8 (0.9)	3.8 (3.7)	-4.3 (-3.3)

Natural resources

In models 2, 3, and 4, Agri decreases productivity via R&D intensity, education level and production structure when Prod90 is below the 67th – 76th percentile. On the other hand, when Prod90 is below 67th – 76th percentile, is the direct impact of Agri on productivity positive. Otherwise it is negative. The end result is, however, that Agri lifts labour productivity with all values of Prod90 in the concerning models. [23] In model 1, this is valid only when Prod90 is relatively low.

Regarding other natural resource variables, forest area seems to lower productivity when Prod90 is below the median level. This is an outcome of the forest variable's strong negative effect on R&D intensity. In contrast, Forest's positive total impact on productivity (when Prod90 is above median level) arises via its effect on production structure. Gas and oil reserves reduce R&D intensity but promote education; as a result, the total impact on productivity is near zero or slightly negative.[24] The only natural resource variable that considerably supports productivity is Minerals.[25] This positivity is explained by the tendency of Minerals to raise R&D intensity and to increase the share of investment and intermediate goods production.

6. Robustness and some other variables

6.1. Random effects estimates

To consider the robustness of the results, we also estimated OLS and random effects panel models. In the OLS models, the dependent variables are the time averages of endogenous variables. Table 7 reports random effects estimates of the panel model, which also includes variables that we regard as endogenous. In the random effects model, R&D intensity and production structure still strongly impact productivity. Education still influences R&D intensity and production structure. The total impact of density (Dens2) on productivity is positive and as strong as its impact in the Hausman-Taylor model. Overall, the impacts of the exogenous variables are quite similar to their respective impacts in the Hausman-Taylor models.

[23] This result may partly be explained by the exclusion of underdeveloped countries from the data.

[24] Overall, these findings are supported by the results previously obtained by Sachs and Warner (1995, 1999 and 2001). Specifically, they found that the export intensity of primary goods (or natural resources) has a negative impact on GDP growth.

[25] Esterline and Levine (2002) obtained a similar result, finding that their crop/minerals variable had a positive impact on GDP. However, Sachs and Warner (1995) obtained the opposite result; according to them, the share of minerals production negatively impacts GDP growth.

Table 7, Random effects estimates of the model in the 4th column in Tables 2 - 5, dependent variables in the columns above

	(1) Prod	(2) R&D-intensity	(3) Share of c-goods	(4) Education
Time-varying endogenous variables				
R&D-intensity	0.160*	-	-0.086**	0.118**
	(0.062)		(0.033)	(0.038)
Share of c-goods	-1.420**	-0.755*	-	-0.642**
	(0.184)	(0.351)		(0.096)
Education	0.078	0.699**	-0.424**	-
	(0.117)	(0.222)	(0.070)	
Time-invariant exogenous variables				
Agri	0.279**	-0.211	0.008	-0.042
	(0.068)	(0.131)	(0.034)	(0.025)
Gas&Oil	-0.003	-0.064**	0.004	0.022**
	(0.011)	(0.017)	(0.003)	(0.005)
Forest	-0.050´	-0.088	-0.022	0.003
	(0.030)	(0.064)	(0.018)	(0.019)
Minerals	0.027	0.094*	-0.022**	-0.045**
	(0.021)	(0.039)	(0.008)	(0.010)
Population	0.001	-0.116	0.019	-0.008
	(0.057)	(0.086)	(0.018)	(0.021)
Dens2	-0.107	0.360*	-0.057´	-0.084*
	(0.076)	(0.144)	(0.029)	(0.036)
Extdens3	-0.166**	-0.264´	-0.054	0.074*
	(0.059)	(0.151)	(0.037)	(0.034)
Agri*Prod90	-0.032**	0.023	-0.005	-
	(0.007)	(0.015)	(0.003)	
Gas&Oil*Prod90	0.002*	-	-	-
	(0.001)			
Forest*Prod90	-	-	-0.006**	-
			(0.002)	
Minerals*Prod90	-0.004*	0.009*	-0.001	-0.002*
	(0.002)	(0.004)	(0.001)	(0.001)
Pop*Prod90	-	-0.010*	0.001´	0.002**
		(0.004)	(0.001)	(0.001)
Dens2*Prod90	0.019**	-	-	-
	(0.006)			
Extdens3*Prod90	0.013*	-	-0.006*	-0.012**
	(0.006)		(0.003)	(0.003)
Prod90	0.030**	0.104**	0.003	-0.009*
	(0.007)	(0.011)	(0.004)	(0.005)
N	714	714	714	714

Arrellano (1987) type robust standard errors in parentheses.

Significance levels: . <0.1 * <0.05 **<0.01.

6.2. OLS model

In the Appendixes, Table A3 reports the OLS estimates obtained from the pooled data. To compare the OLS estimates with the Hausman-Taylor estimates, which are reported in Tables 2 – 5, we have included endogenous regressors in the OLS models. These models are also estimated without these variables; in this case, we use only the time averages of exogenous variables. In OLS – with pooled data - the originally endogenous variables are regarded as exogenous owing to which, the coefficients of endogenous regressors differ remarkably from the respective coefficients in the Hausman-Taylor models. The estimated impact of the exogenous variables is then also different from the results obtained using the Hausman-Taylor estimator. Omitting endogenous regressors, the estimated impacts of exogenous variables, in a way, include the indirect impact of omitting variables. Then, the results become in many respects close to the calculated total impacts described in the previous chapter. In the OLS models with endogenous regressors, the impacts of exogenous geographic and natural resource variables are distinctive in the following ways:

- The direct impact of density (Dens2) still increases R&D expenditure and lowers the share of consumption goods production. The negative direct impact of Dens2 on education level is diluted in the OLS model.
- In the OLS model, the direct net impact of centrality (Extdens3) is no longer significantly different from zero in the R&D equation only. Centrality's direct negative impact on productivity is also weakened in the OLS model.
- In the OLS model, the direct net impact of population on R&D is still negative when Prod90 is above the median level. In OLS, population size also increases the share of consumption goods production, unlike in the Hausman-Taylor models.
- In the OLS model, the direct and total impacts of natural resource variables are, however, in line with the corresponding impacts obtained in the Hausman-Taylor models.

6.3. Impacts on GDP per capita

Most empirical studies on the economics of geography consider the impacts on GDP per capita rather than on productivity. One reason for this choice is the availability of data. Moreover, GDP per capita as a concept differs from labour productivity because it also takes into account the economy's ability to employ people. In the country set we considered, certain countries – including France, Germany, Belgium, Italy and Turkey – drop in rank when we consider GDP per capita instead of productivity. Conversely, countries such as Switzerland, Canada, Ireland and Iceland rise in rank.

Table 8, Dependent variable GDP per capita

	(1) GDP per capita	(2) R&D-intensity	(3) Share of c-goods	(4) Education
Time-varying endogenous variables				
R&D-intensity	0.142*	-	-0.096**	0.134**
	(0.067)		(0.036)	(0.047)
Share of c-goods	-1.567**	-0.756*	-	-0.635**
	(0.185)	(0.354)		(0.095)
Education	**0.162**	0.693**	-0.420**	-
	(0.117)	(0.224)	(0.075)	
Time-invariant exogenous variables				
Agri	0.243*	-0.071	0.032	-0.024
	(0.096)	(0.096)	(0.039)	(0.026)
Gas&Oil	0.002	-0.071**	0.002	0.024**
	(0.014)	(0.020)	(0.005)	(0.006)
Forest	-0.077	-0.126	-0.029	0.010
	(0.060)	(0.082)	(0.026)	(0.021)
Minerals	0.037´	0.156**	-0.013	-0.053**
	(0.022)	(0.052)	(0.009)	(0.014)
Population	-0.042	-0.125	0.004	-0.007
	(0.048)	(0.101)	(0.022)	(0.025)
Dens2	-0.001	0.418*	-0.043	-0.103*
	(0.065)	(0.167)	(0.036)	(0.042)
Extdens3	-0.184**	-0.187	-0.044	0.059´
	(0.070)	(0.136)	(0.043)	(0.032)
Agri*GDPpop90	-0.030**	-	-0.009´	-
	(0.010)		(0.005)	
Gas&Oil*GDPpop90	0.003´	0.009*	0.001*	-0.001
	(0.002)	(0.004)	(0.001)	(0.001)
Forest*GDPpop90	-0.008	-	-0.004*	-
	(0.005)		(0.002)	
Minerals*GDPpop90	-	0.012´	-0.002´	-0.003*
		(0.007)	(0.001)	(0.001)
Pop*GDPpop90	-	-0.037**	-	0.008´
		(0.013)		(0.004)
Dens2*GDPpop90	0.017**	0.054*	0.004*	**-0.010**
	(0.005)	(0.021)	(0.002)	**(0.006)**
Extdens3*GDPpop90	0.023**	-	-0.005	-0.012**
	(0.008)		(0.004)	(0.004)
GDPpop90	0.040**	0.181**	0.009*	-0.024´
	(0.010)	(0.029)	(0.004)	(0.012)
N	714	714	714	714

Arrellano (1987) type robust standard errors in parentheses.
Significance levels: . <0.1 * <0.05 **<0.01.
Bolded values indicate that coefficients are significant at least at the 0.1 level with basic HT standards errors.

The results may change for two reasons; first, because the productivity variable is replaced by the GDP per capita variable and second, because the interactive variable Prod90 is replaced by GDP per capita in 1990, GDPpop90. The direct impact of geography and natural resources are in almost all respects close to the estimates obtained from the model in which productivity was the dependent variable. Nonetheless, the results (obtained from model 4 in Tables 2 - 5) have changed in the following respects:

- The direct net impact of internal density (Dens2) on GDP is substantially positive for all countries, whereas its direct impact on productivity was close to zero or negative for most countries.
- The population size increases the GDP per capita of the countries whose GDP in 1990 was above the median level. In the productivity models, this was not found.
- Regarding natural resource variables, the replacement with GDP causes only minor changes.

6.4. The impacts of certain other variables

In appendix (in Table A4), we present estimation results from model 4 after having added the military expenses per capita (Milit). This variable is treated as exogenous, although on some grounds, one could regard it as endogenous. In any case, Milit lowers R&D intensity and education level. It looks as if the more the government spends on military, the less is left for schooling. Via its impacts on R&D intensity and Education, Milit has a strong negative influence on productivity. The inclusion of Milit also weakens the impact of internal density in all equations. Another additional variable is university ranking index (Univ), which is also treated as exogenous. The university ranking negatively affects education and positively affects R&D intensity. This result may refer to the existence of such a mechanism that is described in chapter A2.3. According to this, investments in the amenity (university) to attract high skills may lower the social returns to school expenditures, which reduces public expenditure on schooling and lowers the education level of ordinary people. That the inclusion of Univ so clearly weakens the impacts of geographical variables on R&D intensity and education, however, indicates that university ranking itself is partly determined by geography. Therefore, it is, after all, highly critical to regard this variable as exogenous.

Replacing the education variable, which take into account PIAAC levels, with an educational attainment variable, Lsecter, in model 4, the negative effect of internal density (Dens2) on education level becomes statistically insignificant. At the same time, the impact of population on education turns negative for the most of the countries. Otherwise, estimates do not change much. These results are not reported.

The urbanization variable (Urb) describes the population concentration in cities with over 300 000 inhabitants and, for certain small countries, the share of the population in the capital. Therefore, it is possible that in an urban country, there are long distances between cities, whereas in a more rural country, population density is high, and distances between the cities and living areas are short, on average. Thus, in the data, countries can be highly urbanized even if density is low. In the model (in the 4th column in Tables 2 – 5), the replacement of the Dens2 variable with the urbanization variable changes the results in the following respects:[26]

- In the education equation, the Urb variable has no effect, whereas the respective impact of the Dens2 variable was remarkably negative.

- The direct net impact of urbanization on R&D intensity is remarkably smaller than the corresponding impact of the Dens2 variable. Urbanization also increases the share of consumer goods, unlike Dens2. For these reasons, the total impact of urbanization on productivity is negative for countries with Prod90 above the 38th percentile, whereas the impact of the Dens2 variable is negative only for the countries whose Prod90 is below the 33rd percentile.

In our results, the positive externalities of urbanisation, which are rather strong at the lower level of development, become dominated by congestion diseconomies when the country comes closer to the technological frontier. These results are consistent with the results previously obtained by Hendersson (2003) and Brülhart and Sbergami (2009).[27]

7. Conclusions

We in this study examine the role of geographical factors for the growth among mainly developed countries. Geography in this study is defined to emphasize the differences between these countries. We thus focus on population density, centrality with respect to neighbours, and population size. In the specification of available natural resources, which are the share of agricultural land area, the share of forest area, gas and oil reserves and minerals production, we also tend to capture meaningful differences among the developed countries.

This study regards geography and natural resources as types of natural advantages that have the potential to increase productivity or GDP per capita. The similarity of the countries in the sample with respect to productivity raises the question whether repercussions exist that dilute the positive impacts of the abovementioned advantages. The mechanism that could create these repercussions is examined in the theoretical part of this study.

Empirically investigating the impacts in the area of interest, we also delve more deeply into the intermediation mechanism so that we can separate direct impacts from indirect impacts. The direct impacts of geography and natural resources can be important, although their total impact is marginal. In fact, we

[26] These results are not reported, either.
[27] This also contradicts Williamson (1965), who concluded that at the initial level of development, agglomerations and related human capital accumulation speeds growth. Later, congestion diseconomies start to retard growth, so that above a certain threshold, a country will not benefit from agglomeration.

discovered that internal density (Dens2) has a negative direct impact on productivity for a country whose Prod90 is at the median level. However, because Dens2 increases R&D intensity, it also has a positive indirect impact, which dominates the total impact. By restricting one's attention to the total impact – which is close to zero in some cases – one loses important information about the intermediating mechanism.

We thus investigated the impacts of geography and natural resources in the framework in which these factors directly impact labour productivity (or GDP per capita) or work through endogenous intermediating variables, namely, R&D intensity, industrial structure (the share of consumer goods production) and the level of adult skills (education). The specific exogenous variables representing geography or natural resources impact each endogenous variable directly; through R&D intensity, industrial structure and education, they also impact the final outcome variable, which is labour productivity or GDP per capita.

It turned out that the total impact of the share of agricultural land area on productivity is slightly positive in most models despite its negative impact on R&D intensity. Forest area and gas reserves, in their entirety, negatively affect R&D intensity and productivity, although forest area supports productivity via its impact on the production structure and gas and oil reserves via their impact on education. Minerals (which are measured in terms of their production) have a positive impact on R&D intensity and on the share of intermediate and investment goods production in total manufacturing, and so Minerals also raise productivity.

The direct productivity impact of the population size, given the other endogenous variables, is negative or close to zero for the median country. In most models, population size has also a negative total impact on productivity or GDP per capita. This is explained by the fact that population size, by its direct impact, decreases both R&D intensity and education level. The magnitude of the negative impact of population on productivity is weakened when the variable that describes average distances within the country (Dens2) is replaced by the traditional population density variable (Dens1).

Our results indicate that high internal population density (Dens 1 or Dens2) increases R&D intensity, which also results in greater productivity. This robust result reflects the influence of agglomeration forces. The total impact is enhanced by the tendency of density to lower the share of consumer goods production. Nonetheless, it is also worth noting that the positive total impact of internal density on productivity is reduced by density's direct negative impact on education. The negative direct impact of density on education may reflect political actions to compensate for the disadvantages of low density. More specifically, it may be related to the impacts of "brain drain" because both outflow (emigration) and inflow (immigration) lower the social returns on education relative to its private returns. Insofar as densely populated countries attract more talent compared to other countries in the sample – without remarkable differences in emigration rates – the incentive to invest in education is weaker in densely populated countries. It is also possible that countries that lack the positive externalities of density act strategically and assume that in the long term, investments in education are paid back in the form of higher productivity. The question is why only sparsely populated countries would take advantage of this strategy. Clearly, this topic needs further research.

There is no clear explanation for why centrality tends to decrease R&D investments. It is possible that remote countries do not attract foreign investments and foreign knowledge capital, which forces them to compensate by self-making. According to this theory, firms would invest in R&D, and public authorities would support these investments to compensate for the lack of foreign investments. Centrality also tends to increase the share of goods other than consumption goods in total manufacturing production, which may be a sign of HME. In most models, the direct impact of centrality on education level is negative among the countries with high Prod90. It is possible that this impact reflects a type of compensation mechanism similar to the one found to characterize the impact of internal density on adult skills. Overall, the total impact of centrality is mostly negative. It is, however, remarkable that the total impact of centrality on productivity increases with the initial level of productivity (Prod90).

Literature

Acemoglu, D., Johnson, S. and Robinson, J. (2001), The colonial origins of comparative development: an empirical investigation. *The American Economic Review* 91(5), 1369–1401.

Acemoglu, D., Johnson, S. and Robinson, J. (2002), Reversal of fortune: geography and institutions in the making of the modern world income distribution. *The Quarterly Journal of Economics*, 117(4), 1231–1294.

Acemoglu D., Aghion P. and Zilibotti F. (2006), Distance to frontier, selection and economic growth. *Journal of the European Economic Association,* 4(1), 37–74.

Aghion, P. and Howitt, P. (1992), A model of growth through creative destruction. *Econometrica*, 60(2), 323 – 351.

Aghion, P. and Howitt, P. (2009), *The economics of growth*. Cambridge: The MIT Press.

Barro, R. and Lee, J. (2013), A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104, 184–198.

Amemiya, T. and MaCurdy, T. (1986), Instrumental-Variable Estimation of an Error-Components Model. *Econometrica*, 54(4), 869–880.

Arellano, M. (1987), Computing Robust Standard Errors for Within-Groups Estimators. *Oxford Bulletin of Economics and Statistics,* 49(4), 431–434.

Barro, R. (2015), Convergence and Modernisation. *The Economic Journal*, 125(585), 911–942.

Behrens, K., Lamorgese, A., Ottaviano, G. and Tabuchi, T. (2004), Testing ‘home market effect’ in a multi-country world: A theory-based approach, CEPR Discussion Paper No. 4468.

Bernard, A., Redding, S. and Schott, P. (2007), Comparative advantage and heterogeneous firms, *Review of Economic Studies*, 74(1), 31–66.

Bernhardt, I. and Jung, B. S. (1979), The Interpretation of Least Squares Regression With Interaction or Polynomial Terms. *The Review of Economics and Statistics,* 61(3), 481–483.

Bertoli, S, Brucker, H., Facchini, G, Mayda, A.-M. and Peri, G. (2011), The battle for brains: how to attract talent. In: T. Boeri, H. Bruecker, F. Docquier, H. Rapoport (Eds.), Brain Gain and Brain Drain, Oxford University Press, Oxford.

Breusch, T., Mizon, G. and Schmidt, P. (1989), Efficient Estimation Using Panel Data. *Econometrica*, 57(3), 695–700.

Breusch, T.; Ward, M. B.; Nguyen, H. T. M. & Kompas, T. (2011a) FEVD: Just IV or Just Mistaken? *Political Analysis,* 19(2), 165–169.

Breusch, T.; Ward, M. B.; Nguyen, H. T. M. & Kompas, T. (2011b) On the Fixed-Effects Vector Decomposition. *Political Analysis,* 19(2), 123–134.

Boschini, A., Petterson, J. and Roine, J. (2007), Resource curse or not: a question of appropriability. *Scandinavian Journal of Economics*, 109(3), 593–617.

Boschini, A., Petterson, J. and Roine, J. (2013), The resource curse and its potential reversal. *World development*, 43, 19–41.

Brülhart, M. and Sbergami, F. (2009), Agglomeration and growth: cross-country evidence. *Journal of Urban Economis*, 65(1), 48–63.

Constantini, V. and Paglialunga, E. (2014), Elasticity of substitution in capital-energy relationships: how central is a sector-based panel estimation approach? SEEDs Working Paper Series, 13/2014.

Crozet, M. and Trionfetti, F. (2008), Trade costs and the home markets effect. *Journal of International Economics.* 76(2), 309–321.

Davis, D. and Weinstein, D. (1996), Does economic geography matter for international specialization? NBER Working Paper No. 5706.

Davis, D. and Weinstein, D. (2003), Market access, economic geography and comparative advantage: An empirical test. *Journal of International Economics,* 59(1), 1–23.

Deardorff, A. (2001), Fragmentation in simple trade models. *The North American Journal of Economics and Finance,* 12(2), 121–137.

Delgado, M., Porter, M. E. and Stern, S. (2014), Clusters, convergence, and economic performance. *Research Policy*, 43(10), 1785–1799.

Easterly, W. and Levine, R. (2003), Tropics, germs, and crops: how endowments influence economic development. *Journal of Monetary Economics,* 50(1), 3–39.

Feenstra, R., Inklaar, R. and Timmer, M. (2015), The Next Generation of the Penn World Table. *American Economic Review*, 105(10), 3150–3182.

Gallup, J. L.; Sachs, J. D. and Mellinger, A. D. (1999), Geography and Economic Development. *International Regional Science Review*, 22(2), 179–232.

Glaeser, E., La Porta, R., Lopez-De-Silanes, F. and Shleifer, A. (2004), Do institutions cause growth? *Journal of Economic Growth,* 9(3), 271–303.

Greene, W. (2011) Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models? *Political Analysis,* 19(2), 135–146.

Grünfeld L. (2002), International R&D spillovers and the effect of absorptive capacity: an empirical study, NUPI (Norwegian Institute for International Affairs), mimeo.

Hanson G. H. and Xiang C., (2004), The Home-Market Effect and Bilateral Trade Patterns. *American Economic Review*, 94(4), pages 1108–1129.

Havranek, T., Horvath, R. and Zeynalov, A. (2016), Natural Resources and Economic Growth: A Meta-Analysis. *World Development*, 88, 134–151.

Hausman, J. A. (1978), Specification Tests in Econometrics. *Econometrica,* 46(6), 1251–1271.

Hausman, J. and Taylor, W. (1981), Panel data and unobservable individual effects. *Econometrica,* 49(6), 1377–1398.

Head, K. and Mayer, T. (2002), Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade. *Working Paper 2002-01, CEPII research center*.

Helpman, E. (1981), International trade in the presence of product differentiation, economies of scale, and monopolistic competition: A Chamberlin-Heckscher-Ohlin approach. *Journal of International Economics* 11(3), 305–340.

Helpman, E. (1987), Imperfect competition and international trade: evidence from fourteen industrial countries. *Journal of The Japanese and International Economics*, 1(1), 62–81.

Helpman, E. and Krugman, P. (1985), Market Structure and foreign Trade, Cambridge, MA, MIT Press.

Henderson, V. (2003), The urbanization process and economic growth: the so-what question. *Journal of Economic Growth*, 8(1), 47–71.

Hernesniemi, H. (2000), Evaluation of Estonian innovation system. *The Research Institute of the Finnish Economy, Discussion papers* 772, 1–68.

Jaffe A., Trajtenberg, M. and Henderson R. (1993), Geographical localization of knowledge spillovers as evidence by patent citations. *Quarterly Journal of* Economics, 108(3), 577–598.

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

Koesler, S. and Schymura, M. (2015), Substitution elasticities in a constant elasticity of substitution framework – empirical estimates using nonlinear least squares. *Economic Systems Research* 27(1), 101–121.

Krugman, P. (1980), Scale economies, product differentiation, and the patterns of trade. *The American Economic Review*, 70(5), 950–959.

Krugman, P. (2009), Increasing returns and comparative advantage world, in *Comparative advantage, economic growth, the gains from trade and globalization, and computational trade modeling: Festschrift Papers in Honor of Alan V. Deardorff.* IPC Working Paper Series number 91.

Lederman, D. and Saenz, L. (2005), Innovation and development around the World, 1960-2000. Policy Research Working Paper No. 3774, World Bank.

Lehto, E. (2007), Regional impact of research and development on productivity. *Regional Studies,* 41(5), 623–638.

Lehto, E., Böckerman, P. and Huovari, J. (2010), The return to the technological frontier: The conditional effect of R&D on plant productivity in Finnish manufacturing. *Papers in Regional Science,* 90(1), 91–109.

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

Maurseth P. and Verspagen B. (2002), Knowledge spillovers in Europe: a patent citations analysis, Scandinavian Journal of Economics 104(4), 531–545.

Mayer, T. and Zignago, S. (2011), Notes on CEPII's distance measures: the GeoDist database. *CEPII Working Paper*.

Moretti, E. (2004), Estimating the social returns to higher education: evidence from longitudal and repeated cross-sectional data. *Journal of Econometrics,* 121(1–2), 175–212.

Moretti, E. (2006), Private and social returns to education, mimeo, Department of Economics, University of Berkley.

OECD (2016). Input-output tables. Organisaton for Economic Co-operation and Development, OECD Data.

Pesaran, M. Hashem (2006), Estimation and inference in large heterogeneous panels with a multi-factor error structure. *Econometrica,* 74(4), 967–1012.

Plümper, T. and Troeger, V. (2007), Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis* 15(2), 124–139.

Rauch, J. E. (1999), Networks versus markets in international trade. *Journal of International Economics*, 48(1), 7–35.

Rebernik, D. (2014), Population and spatial development of settlements in Ljubljana urban region after 2002. *Dela,* 42, 75–93.

Reichl, C., Schatz, M. and Zsak, G. (2016), World-Mining-Data 31.

Roback, J. (1988), Wages, rents and amenities: differences among workers and regions, Economic Inquiry no. 261, 23–41.

Rodrik, D., Subramanian, A. and Trebbi, F. (2004), Institutions rule: the primacy of institutions over integration and geography in economic development. *Journal of Economic Growth*, 9(2), 131–165.

Romer, P. (1990), Endogenous technological change. *Journal of Political Economy,* 98(5), 71–102.

Sachs, J. and Warner, A. (1995), Natural resource abundance and economic growth, NBER Working Paper Series No. 5398.

Sachs, J. and Warner, A. (1999), The big push, natural resource booms and growth*. Journal of Development Economics*, 59(1), 43–76.

Sachs, J. and Warner, A. (2001), The curse of natural resources. *European Economic Review*, 45, 827–838.

Serlenga, L. and Shin, Y. (2007), Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics,* 22(2), 361–381.

Spence, M. (1973), Job Market Signaling, *The Quarterly Journal of Economics, 87(3),* 355–374.

Timmer, M., Dietzenbacher, E., Los, B., Stehrer, R. and de Vries, G. (2015), An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics,* 23(3), 575–605.

Torres, N. and Afonso, O. (2008), Re-evaluating the impact of natural resources on economic growth, FEP Working Papers No. 278.

Vandenbussche J, Aghion P, Meghir C (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth,* 11(2), 97–127.

Vieira, F., MacDonald, R. and Damasceno, A. (2012), The role of institutions in cross section income and panel data growth models: A deeper investigation on the weakness and proliferation of instruments. *Journal of Comparative Economics,* 40(1), 127–140.

Williamson, J. (1965), Regional inequality and the process of national development. *Economic Development and Cultural Change,* 13(4), 3–45.

World Bank (2016), The World Bank DataBank — Explore. Create. Share.

World Trade Report (2008), World Trade Organization.

Appendix 1 - the construction of the data

Labour productivity

Productivity level (prod). Real PPP-adjusted GDP divided by average hours multiplied by employment, i.e., labour productivity. The variables used are Rgdpe, Emp and Avh from PWT9.0. Unit: 2011 dollars. Source: Penn World Tables 9.0 (PWT9.0), (Feenstra et al., 2015).

GDP per capita

GDP per capita is calculated using real PPP-adjusted GDP (Rgdpe) from PWT9.0 and population data from World development indicators. As Feenstra et al. (2015) note, the expenditure side of purchasing power parity- (PPP-) fixed GDP is useful when comparing living standards across countries. Unit: 2011 dollars. Source: Penn World Tables 9.0 (PWT9.0), (Feenstra et al., 2015), WDI.

R&D intensity

To measure R&D intensity we use real R&D expenditure divided by real GDP. The R&D expenditure time-series is constructed using several sources, but the main source of the data the OECD database (for the following countries: ARG, AUS, AUT, BEL, CAN, CHE, CHL, CHN, CZE, DEU, DNK, ESP, EST, FIN, FRA, GBR, GRC, HUN, IRL, ISL, ISR, ITA, JPN, KOR, MEX, NLD, NOR, NZL, POL, PRT, ROU, RUS, SWE, SVK, SVN, TUR and USA). For certain countries, we fill in missing information using data from Eurostat (BGR, LTU, LVA). In addition, three values for EST (1995-1997) are from Hernesniemi (2000). We also gathered data for other countries. For BRA, THA and CHL the figures are from World Bank data and from UNESCO's statistical yearbooks (CHL 1995,1996 and BRA 1995,1996). One value for BRA (1998) is from Lederman & Saenz (2005). Some of these additional figures were represented in the form of nominal R&D expenditure as a share of nominal GDP. Therefore, these were first multiplied by nominal GDP values to obtain the total value of R&D expenditure which was then deflated by the price level of gross fixed capital formation. This deflator is from the OECD database and the AMECO database (CHL, ISR, KOR, RUS, SVK, SVN). When this deflator was not available, we used the GDP deflator (WB ARG, WB BRA, OECD LVA, WB THA). Finally, to obtain a balanced panel for R&D intensity, the following values are interpolated: AUS: 1995, 1997, 1999, 2001, 2003, 2005, 2007, 2009 BRA: 1997, 1999 CHE: 1995, 1997, 1998, 1999, 2001-2003, 2005, 2006, 2007 CHL: 1997-2006 GRC: 1996, 1998, 2000, 2002 ISL: 1996, 2004, 2010 NOR: 1996, 1998, 2000 NZL: 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010 SWE: 1996, 1998, 2000, 2002 THA: 1998, 2008, 2010. In addition, three values are extrapolated: ARG: 1995 and THA: 1995, CHE: 2011. Unit: % of GDP. Source: PWT9.0, OECD, Eurostat, AMECO, WDI.

Higher educational attainment

To construct a measure that represents higher education attainment, we sum the upper secondary and tertiary attainment shares in the age group from 25 to 64. This variable is called Secter. By the Eurostat's definition, our measure includes educational attainment levels 3-8. According to this, we exclude less than primary, primary and lower secondary education (0-2). Eurostat provides data for 28 countries of our study. Some observations for these countries are still are missing. In this part, we supplement our data using the information provided by the OECD. Additionally, for the rest of the countries (USA, CAN, KOR, AUS, BRA, ARG, CHL, CHN, ISR, MEX and NZL), we use data provided by the OECD. The values for THA (2004, 2006, 2010) are from the World Bank. The time series for BRA is also completed using World Bank data. For RUS and JPN, we construct time series using enrolment and some values for educational attainment. The data for RUS are obtained from the OECD, World Bank and UNESCO, and those for JPN are from Population Statistics of Japan 2017 and World Bank data. Some values are still missing for the rest of the countries. The missing values of the years 1995, 2000 and 2005 are obtained from Barro-Lee's (2013) dataset. From this data, which also include the lower secondary level of education, we use only the information contained in time variation to make an assessment of the educational attainment of higher education that corresponds to OECD data. After this, we use linear interpolation to fil in remaining missing values. For CHN and THA, the year 2011 is linearly extrapolated. Unit: % of population aged 25 to 64. Source: Eurostat, OECD, WDI, Barro-Lee (Barro and Lee, 2013).

PIAAC levels

Differences between countries in education level are assessed using the OECD's adult skill surveys. For each country for which the data are available, we construct an average (over time) level of adult skills using PIAAC scores and IALS literacy section scores. These data are extracted from OECD data and reports (Data on survey of adult skills (PIAAC); Adult literacy in OECD countries: technical report on the first international adult literacy survey, 1997; Literacy in the information age - final report of the international adult literacy survey, 2000; Learning a living - first results of the adult literacy and life skill survey, 2005; OECD Adult skills outlook 2013 – first results from the surveys on adult skills; OECD skill studies – the survey of adult skills, 2016). PIAAC and IALS scores combined cover only 32 countries out of the 42 countries in our data. For the rest of the countries, we estimate missing values using PISA results. The PISA data are obtained from the OECD PISA reports. We first construct a representation of the overall PISA performance. This is the average score in PISA tests in the years 2000, 2003, 2006 and 2009 in the scientific, mathematical and literary sections.

China is the only country that lacks both PIAAC and PISA scores. The PISA results concerning Shanghai are not considered as representative. The missing overall PISA value for China is estimated using the following OLS regression (with standard deviation in parenthesis):

$$Lpisa = \frac{5.89}{(0.13)} + \frac{0.09}{(0.03)} * Lsecter - \frac{0.07}{(0.05)} * Lgini - \frac{0.06}{(0.02)} * Soc + \frac{0.26}{(0.10)} * lLabsh,$$

where $\mathrm{Lsecter}$ is the percentage of the population with higher educational attainment; Llabsh is the labour share (the share of labour costs of GNI) in PWT9.0; and Lgini is the Gini coefficient obtained from the WDI. These variables are in logs of the means over time so that there are 41 observations in total. Additionally, Soc is a dummy for a former or current socialist country.
Then, we estimate the missing PIAAC values (Lpiaac) for ten countries (ARG, BGR, BRA, CHN, ISL, LVA, MEX, ROU, THA and TUR) using the following OLS regression:

$$Lpiaac = \frac{0.74}{(0.07)} + \frac{0.70}{(0.12)} * Lpisa + \frac{0.11}{(0.02)} * Lsecter + \frac{0.02}{(0.01)} * Col14,$$

where the number of observations is 32, and $\mathrm{Col14}$ is a dummy for colonial powers in year 1914. Unit: Index. Source: WDI, OECD, PWT9.0.

Secondary plus tertiary educational attainment rates adjusted using PIAAC levels

To reach our education variable (Education), we combine the information in the variable PIAAC with information in the Secter variable. This is done by dividing Secter for each country by its 2011 value. The result is then multiplied by PIAAC. Intuitively, we adjust the higher educational attainment by the quality (as represented by PIAAC) of education in each country. The correlation between the logarithms of higher educational attainment variable, Secter, and our quality adjusted education level variable, Education, is 0.89. Unit: Index. Source: Eurostat, Barro-Lee, WDI, OECD.

Total mineral production divided by population

The Minerals variable is obtained by dividing total mineral production (not including diamonds and mineral fuels) in 2014 by population in 2011. For countries with no mineral production (ISL), the values are set to an epsilon to enable logarithmic transformation. This is the only exogenous variable of which we could not find past values (from years 1990–1994). Unit: Million USD. Source: World Mining Data, Reichl et al. (2016).

Percentage of land area covered by arable agricultural land

Agricultural land (Agri) refers to the share of total land area that is arable, under permanent crops, and under permanent pastures. The data are mainly from the WDI. For some countries (BEL, CZE, EST, LVA, LTU, SVK, SVN), we supplement WDI data using the information provided by the OECD to fill in the missing values. The average of this measure over the period 1990–1994 is used in this study. Unit: % of land area. Source: WDI OECD.

Percentage of land area covered by forest

This variable is the share of total land area that is covered by forest (Forest). The average share over the period 1990–1994 is used in this study. Unit: % of land area. Source: WDI.

Military expenditure divided by GDP

Data for the military expenditure share of GDP (Milit) are from the Stockholm International Peace Research Institute (SIPRI). For ISL, the years before 2009 are unavailable. For practical reasons, we set the value for ISL for the years 1995–2008 at 0.25, although the true value is possibly closer to zero or even zero. Unit: % of GDP. Source: SIPRI.

Natural gas and crude oil reserves divided by population

This variable is a combination of proven reserves of natural gas (Gas&Oil) in 1990 (trillion cubic feet) and proven reserves of crude oil (billion barrels). We combine these by converting both into monetary values and then summing these values together. This is done by using the following prices, which are the average prices from the period 1990-2012: $3 per thousand cubic feet and $50 per barrel. This aggregate is then divided by the average population from 1990-1994. The data are from the Energy Information Administration (EIA). For countries with no reserves (BEL, CHE, EST, FIN, KOR, ISL, LVA, PRT, SWE, SVN), the values are set to an epsilon to enable logarithmic transformation. If information for 1990 is not available, the earliest available value is used instead. Unit: Index. Source: EIA.

Geographical density and centrality

To measure the effects of internal distance and geographic centrality (inverse of remoteness) we construct a centrality measure using the distances provided by the GeoDist database of Mayer and Zignago (2011).
We construct our geographical centrality measure as follows:

$$centrality_i = \frac{M_i}{d_{ii}} + \sum_{k \neq i}^{N} \frac{M_k}{d_{ik}} \qquad (a1)$$

where M denotes an "economic" mass variable; d is the distance measure; and k, i denotes the country. In (a1), d_{ii} is calculated as $d_{ii} = 0.67 * \sqrt{\frac{area_i}{\pi}}$, whereas d_{ik} is obtained by inserting the locations of the most important city in the own county (i) and in a neighbouring country (k) into the formula of great circle-distance to achieve the geodesic distance between these two cities. This distance is regarded as the average distance between the two countries to which these two cities belong.

In formula (a1), centrality is divided into two parts: the inverse of the country's own distance (the first term on the right-hand side) and the inverse of the distance to other countries (second term on the right-hand side). In (a1), d_{ii} can be interpreted as being a good approximation for the average distance between any two points uniformly distributed within the area of a disk (representing county's area). Dividing d_{ii} by population, we obtain the inverse of Dens2, which is an approximation for the average distance of persons within a country. Variable Dens2 then describes the inverse of this distance, being a kind of measure for internal density.

In the second term, d_{ik} describes the distance between two sites in any two countries, i and k. Using population as a mass variable in the second term of (a1), we obtain Extdens 1 for external density, a proxy for the inverse of the average distance of population from country i to country k. When GDP is used as a mass variable in the second term of (a1), we obtain the variable Extdens2 for external density. Ideally, our centrality variable would require accurate data for all countries of the world. The World Bank offers population and GDP data for most countries in the world for the period 1995-2011. We assume that this coverage is sufficient to make the measure reliable. Finally, we use the averages (over years 1990–1994) of these measures.

We also modify the Extdens2 variable to better account for the overall area of a country. The intuition is that if a country has a large area, its distance to other countries is larger; however, this distance does not necessarily mean that the country is very remote (consider, for example, Canada next to the USA market), and in fact, Extedens2 correlates negatively with the country's area. Therefore, we make the following adjustment to the measure Extdens2:

$$\mathrm{Extdens2}_k = \alpha + \beta_1 Larea_k + \beta_2 Larea_k^2 + \epsilon$$

where $Larea$ denotes the log of country k's overall area. We then take the residual ϵ from this regression and use it as our area-adjusted centrality variable (Extdens3). By this modification we are able to take into account the presence of borders which Rauch (1999) did by adding an extra dummy variable. The area variable is from the World Bank. Unit: Index. Geodist (Mayer and Zignago, 2011), WDI.

Population density

This variable (Dens1) is the density (average of population (90–94)/average of area (90–94)) of a country. The data are from World Bank. For Belgium, the density time series is continued with OECD density growth rates for the years 1990–1994. The average of this measure over the years 1990–1994 is used. Unit: population per sq. km of land area. Source: WDI, OECD.

Urbanisation

This variable (Urb) is the percentage of the total population residing in each urban agglomeration with 300,000 or more inhabitants in 2014. The data are from the United Nations database. For certain small countries (ISL, EST, LTU, LVA and SVN) that have no cities with over 300,000 inhabitants, we use the share of the total population living in the capital district. These data are from the World Bank (EST, LTU, LVA), Statistics Iceland (ISL) and Rebernik (2014) (SVN). Unit: % of total population. Source: United Nation, WDI, Statistics Iceland, Rebernik (2014).

Consumption goods share of total manufacturing

The measure of consumption goods divided by total manufacturing (Share of c-goods), is constructed as follows. First, we calculate sectoral shares of consumer goods from the WIOD tables, as follows:

$$S_j = \frac{C_j}{C_j + I_j + V_j},$$

where $C_j = \sum_i^N C_{ij}$, $I_j = \sum_i^N I_{ij}$, and $V_j = \sum_i^N V_{ij}$. C_j is the amount of production consumed by the end-user; I_j is the amount of production invested and V_j is the amount of production used as intermediate products; C_j is the sum of the final consumption expenditure by households, final consumption expenditure by non-profit organizations serving households (NPISH), and final consumption expenditure by the government; I_j is the sum of gross fixed capital formation and changes in inventories; and the letter j denotes the manufacturing sector ($N = 13$) and the letter i denotes the country. The sector classifications are 15t16, 17t19, 20, 21t22, 23, 24, 25, 26, 27t28, 29, 30t33, 34t35, and 36t37, in accordance with the OECD's (2016) Input-Output tables. In the OECD's Input-Output tables, the sectors c17, c18 and c19 are reported as a sum (c17t19) and not separately, as they are in the WIOD tables. The weight for c17t19 is therefore chosen as the average weight of 17t18 and 19. That is, when calculating the weights, $N = 14$. Thus, S_j represents the percentage of production in the manufacturing sector j that is consumed as an end-product in the world. Using S_j's as "universal" weights, we next construct a measure that reflects the share of consumer goods in the entire manufacturing sector in country i. For this process, we use data from the OECD's Input-Output tables, which enables greater country coverage. We construct this measure as follows:

$$Q_C^i = \sum_j^N S_j \, Q_j^i \, .$$

$$\text{socgoods}_\text{i} = \frac{Q_C^i}{Q^i}$$

Variable Q_C^i represents the production of consumption goods in country i's manufacturing sector. Q^i represents the total value added in the manufacturing sector in country i. This procedure is repeated for each year separately to construct a time series of this variable. Unit: % of value added in manufacturing sector. Source: WIOD (Timmer et al., 2015), OECD.

Quality of universities

This variable (Univ) is constructed from the data provided by the Center for World University rankings (CWUR) in the year 2015. The variable is constructed as follows:

$$univ = (\sum_{i=1}^{N} \frac{1}{10 + r_i^k})/pop_k$$

where r is country's k university ranking i. We calculate this measure from the 1000 best ranked universities. For LVA, there are no universities within this ranking. For practical reasons, we set one university for LVA, which for the rank is 1000. Unit: Index. Source: CWU.

Appendix 2 – Models and simulations

A2.1. A model for growth and innovative activity

We tend to characterize the possible impacts of geography and natural resources in a growth model that emphasizes the central role of innovative activities. We look at the market where the production of final goods (either consumer or investment) is rather competitive and the intermediate producer has a monopoly. The intermediate monopolist sets its prices to obtain all of the profits in the production chain. Thus, the intermediate good monopolist is also motivated to invest in R&D, which determines the level of technology in the production process that manufactures consumer or investment goods. Because the intermediate good monopolist makes production decisions based on demand for its products in the final stage of production, the intermediate good monopoly can be modelled to manufacture its products using the final good as an input. In this regard, the model follows Aghion and Howitt (2009)[28]. However, we assume that the production of final goods is divided between consumer and investment goods. We also consider potential differences in factor substitutability. Therefore, instead of regarding the production function as Cobb-Douglas, we assume that it is CES. In addition, we assume that intermediate goods in this process are made from natural resources and labour. In our model, the technology in which the intermediate good monopolist invests affects the entire process through which labour and natural resources are transformed into intermediate goods, which

[28] See also Aghion and Howitt (1992).

are then used to make final goods. If the supply of natural resources is abundant, then natural resources are used in primary production processes, where the level of technology is assumed to be lower compared with manufacturing. The geography – i.e., density, centrality and population size – may create natural advantages in consumer goods production or R&D activity. Finally, we discuss the role of human capital and the social motives to invest in education.

The main features of the model are as follows:

- Consumer goods production and investment goods production have their own production functions.
- The division between investment and consumer goods production is determined by the level of technology in each sector. Because these goods are traded globally with fixed global prices, the domestic demand ratio does not fix the respective production ratio.
- Asymmetries (which in part reflect the impact of increasing returns to scale) are built into the production function and the R&D cost function.
- Intermediate goods are manufactured in the first phase using labour and natural resources as inputs. Later, investment goods and consumer goods are manufactured such that for each unit of an intermediate good produced, one unit of the value is added to the final good.

The production function for investment goods is

$$(1)\ Y_I = [\alpha(A_t L_I)^r + (1-\alpha)(x_I)^r]^{\frac{1}{r}}.$$

where

Y_I = the total production of investment goods

A_t = the level of technology in the production of investment goods

L_I = labour in the production of i-goods

x_I = an intermediate good in i-goods production.

In the above equation, $0 < \alpha < 1$ is a factor share parameter. To simulate the impacts of the various parameters of the model, we need to set the elasticity of substitution between L_I and x_I, which is $\frac{1}{1-r}$, on the realistic level somewhere around 0.3.[29] That is the main reason why we use the CES production function instead of Cobb-Douglas, for which the elasticity of substitution is one.

The respective production function for consumer goods production is

[29] The findings of Constantini and Paglialunga (2014) and Koesler and Schymura (2015) support this view.

(2) $Y_c = [\alpha(B_t L_c)^r + (1-\alpha)(x_c)^r]^{\frac{1}{r}}$,

where

Y_C = the total production of consumer goods,

B_t = level of technology in the production of consumer goods,

L_C = labour in the production of consumer goods, and

x_C = intermediate goods in consumer goods production.

The prices of investment and consumer goods are normalized as being one. Following Aghion and Howitt (2009), we assume that for one unit of an intermediate good is needed one unit of final good. The rest of the final good (Y_I and Y_c), which is not used for the intermediate good, is used for R&D investments, fixed investments[30] and consumption. As in Aghion and Howitt (1992), we assume that the intermediate goods firm is a monopoly and that the price p_{xI} is determined based on the value of the marginal product, $\frac{\partial Y_I}{\partial x_I}$, which reflects the inverse demand curve faced by the intermediate goods firm. Thus,

(3) $p_{xI} = [\alpha(A_t L_I)^r + (1-\alpha){x_I}^r]^{\frac{1-r}{r}}(1-\alpha)x_I^{r-1}$

and

(4) $p_{xc} = [\alpha(B_t L_c)^r + (1-\alpha){x_c}^r]^{\frac{1-r}{r}}(1-\alpha)x_c^{r-1}$.

We assume that intermediate and investment goods production is characterized by increasing returns to scale, owing to which, the profit function of these goods includes fixed costs K_I. On balance, in the constant returns to scale sector (consumer goods production), competition is assumed to be hasher what for the final price is diminished by parameter θ (< 1). Thus, the profits of the intermediate goods monopolist from the production of investment goods and consumer goods are $\pi_{xI} = p_{xI}x_I - c_I x_I - K_I$ and $\pi_{xc} = \theta p_{xc}x_c - c_c x_c$, respectively.

As an alternative conjecture, we may think that big home markets favour consumer goods production. Firms supply consumer goods at a uniform price level in all countries. However, citizens may favour products from their own country, which decreases the price elasticity of local goods. Consumer goods markets could therefore be differentiated based on production location, which increases the supply of consumer goods in relation to the size of a country's economy if fixed costs exist and if marginal costs vary within each country according to the specific distribution. A small country would then decrease production of consumer goods

[30] For purposes of simplification, our model does not include fixed capital as an input. Replacing L_I with $L_I^{\beta}K_I^{1-\beta}$ (with K as capital) would not change the results.

through stricter entry conditions. A country's size thus increases the marginal income of consumer goods production, which is taken into account by multiplying p_{xc} by θ (>1) in the profits function. The parameter θ would then increase with population size.

Parameters c_I and c_C describe marginal costs that the intermediate good monopolist regards as given. The intermediate goods monopolist regards also L_I and L_c as given although the use of intermediate goods impact these variable through labour market repercussions. These assumptions can be regarded as describing the bounded rationality of the decision maker. On the other hand, we can consider them as simplification: the second order impacts of marginal magnitude are, for simplicity, ruled out. From the first-order conditions of profit maximization with respect to x_I and x_C, we can then derive expressions

(5) $[\alpha(A_tL_I)^r + (1-\alpha){x_I}^r]^{\frac{1-r}{r}} {x_I}^{r-1}(1-\alpha)\frac{[r\alpha(A_tL_I)^r+(1-\alpha)x_I^r]}{[\alpha(A_tL_I)^r+(1-\alpha){x_I}^r]} - c_I = 0$

and

(6) $\theta[\alpha(B_tL_c)^r + (1-\alpha){x_c}^r]^{\frac{1-r}{r}} {x_c}^{r-1}(1-\alpha)\frac{[r\alpha(B_tL_c)^r+(1-\alpha)(x_c)^r]}{[\alpha(B_tL_c)^r+(1-\alpha){x_c}^r]} - c_c = 0.$

for the optimal levels of intermediate goods. Using (6), we can express maximum profits π_{xI}^* in the form

(7) $\pi_{xI}^* = [\alpha(A_tL_I)^r + (1-\alpha){x_I}^r]^{\frac{1-2r}{r}}(1-\alpha)(1-r)\alpha(A_tL_I)^r{x_I}^r - K_I.$

Similarly,

(8) $\pi_{xc}^* = \theta[\alpha(B_tL_c)^r + (1-\alpha){x_c}^r]^{\frac{1-2r}{r}}(1-\alpha)(1-r)\alpha(B_tL_c)^r{x_c}^r.$

Following Aghion and Howitt (2009), we assume that an innovator activity – for which one entrepreneur is responsible – uses a final output as its only input. As R&D used for innovation increases, the number of these inputs used also increases. It follows that labour inputs L_I and L_c are also used to produce innovations.

As in Aghion and Howitt (2009), success in innovation activity and the assumed technological improvement in the case of success should determine the level of technology involved in the production of an intermediate product. The probability of success μ_I for an innovation that directly benefits intermediate goods production when the final good is an investment good is governed by the equation

(9) $\mu_I = \lambda\left(\frac{D_I}{A_t}\right)^{\sigma},$

where λ is a Schumpeterian efficiency parameter and D_I denotes R&D investments. The optimal level for D_I is obtained by maximizing

(10) $\pi_{A_t} = \lambda\left(\frac{D_I}{A_t}\right)^{\sigma}\pi_{xI}^* - d_ID_I$

with respect to D_I. In (10), d_I describes unit costs. The optimal D_I obtained from the first-order conditions of profit maximization is

(11) $\lambda\sigma D_I^{\sigma-1}\left(\frac{1}{A_t}\right)^{\sigma}\pi_{Ix}^{*}-d_I=0.$

(12) Following Aghion and Howitt (2009), we assume that if an innovation is successful, it increases A_{t-1} by the amount φ, so that $A_t = φA_{t-1}$ and thus

(13) $\frac{A_t-A_{t-1}}{A_{t-1}}=\varphi-1.$

By the law of large number, the long-run economic growth which is the expected value of an increase in productivity, for given φ is then

$$E\left(\frac{A_t-A_{t-1}}{A_{t-1}}\right)=\mu(\varphi-1).$$

Using expression (7) and condition (11), we are able to express $E(A_t)$ in the form

(14) $E(A_t)=A_{t-1}(\varphi-1)\lambda^{\frac{1}{1-\sigma}}\left(\frac{\sigma}{d_I}\right)^{\frac{\sigma}{1-\sigma}}E(A_t)^{-\frac{1}{1-\sigma}}(\pi_{Ix}^{*})^{\frac{\sigma}{1-\sigma}}+A_{t-1}.$

Similarly, $E(B_t)$ is obtained with the equation

(15) $E(B_t)=B_{t-1}(\varphi-1)\lambda^{\frac{1}{1-\sigma}}\left(\frac{\sigma}{d_I}\right)^{\frac{\sigma}{1-\sigma}}E(B_t)^{-\frac{1}{1-\sigma}}(\pi_{cx}^{*})^{\frac{\sigma}{1-\sigma}}+B_{t-1}.$

In (13), d_c is a cost parameter of R&D investments in consumer goods production. We hypothesized above that quality matters more in the sale of investment goods than in the sale of consumer goods. Therefore, we assume that $d_I < d_c$, that is, it is costlier to obtain innovations that enhance sales and profits in the production of consumer goods than in the production of investment goods.

It is assumed that raw materials are potentially fit for both investment goods and consumer goods. These raw materials are traded in the global market or used locally to manufacture intermediate goods. Reserves of natural resources – parts of which are used for annual production - are thus assumed to be of two types. Let R_I^s denote the supply of natural resources to intermediate goods that can potentially be used for investment goods, and let R_I^d denote the respective demand. Let R_c^s and R_c^d denote supply and demand for raw materials ultimately used for consumer goods. Let L_{xI} denote the labour that is allocated to produce x_I. For the sake of simplicity, we assume that a unit of an intermediate good is the sum $x_I = L_{xI} + R_I^D$ and that a unit of an intermediate good uses labour and natural resources inputs in fixed ratios, such that

$$\frac{L_{xI}}{R_I^d}=h_I,$$

where 0 < $h_I < 1$. Then,

(16) $R_I^d = \frac{1}{1+h_I} x_I$

and

(17) $L_{xI} = \frac{h_I}{1+h_I} x_I$.

The total cost of production of x_I is $wL_{xI} + P_{RI}R_I^D$, where w is the uniform wage level and p_{RI} is the effective price of natural resources used to produce intermediate goods for investment goods. Accordingly, unit costs c_I have the equation

(18) $c_I = \frac{h_I}{(1+h_I)} w + \frac{1}{1+h_I} P_{RI}$.

R_c^d, L_{xc} and c_c are obtained by similar equations

(19) $R_c^d = \frac{1}{1+h_c} x_c$,

(20) $L_{xc} = \frac{h_c}{1+h_c} x_c$

and

(21) $c_c = \frac{h_c}{(1+h_c)} w + \frac{1}{1+h_c} P_{Rc}$,

respectively, where h_c is the ratio of labour input to natural resources input, and p_{Rc} is the effective price of raw materials for consumer goods. The effective prices p_{RI} and p_{Rc} are determined by local prices p_{RI}^l and p_{Rc}^l and global prices p_{RI}^t and p_{Rc}^t. Local prices are lower due to savings in transport and other logistics costs. Thus, $p_{RI}^l < p_{RI}^t$ and $p_{Rc}^l < p_{Rc}^t$. When $R_I^s > R_I^d$ and $R_c^s > R_c^d$, all of the raw materials used for intermediate production can be satisfied with local supply at local prices. Then, the portion of supply that exceeds demand is sold in the global market at fixed prices p_{RI}^t and p_{Rc}^t. If $R_I^s < R_I^d$, then the price of the portion of demand that is satisfied by local supply is at the local level, and for the remaining demand, which is $R_I^d - R_I^s$, the price is p_{RI}^t. The same applies to consumer goods. The effective price for raw material in the production of investment goods is then

(22) $p_{RI} = \frac{R_I^s}{R_I^d} p_{RI}^l + \frac{R_I^d - R_I^s}{R_I^d} p_{RI}^t$.

The same is true for price p_{Rc}. When $R_c^s < R_c^d$, p_{Rc} has the equation

(23) $p_{Rc} = \frac{R_c^s}{R_c^d} p_{Rc}^l + \frac{R_c^d - R_c^s}{R_c^d} p_{Rc}^t$.

In the model considered, we abstract from unemployment. The labour supply L is used either for investment goods, consumer goods or raw material (natural resources) production, so that

(24) $L = L_I + L_{xI} + L_c + L_{xc} + L_R.$

In the above equation, L_R is the labour used for raw material production. We assume that

(25) $L_R = s(R_I^s + R_c^s - R_I^d - R_c^d).$

In the situation considered, firms are classified as either final goods producers or intermediate goods producers. At the aggregate level, the nominal value added in manufacturing production is $Y_I + Y_c$, which includes the value added of final goods producers, intermediate goods producers and providers of natural resources for intermediate goods firms. For example, in the investment goods sector, the nominal value added of the final goods producers alone is $Y_I - p_{xI}x_I$. The value added of the intermediate goods monopoly is $p_{xI}x_I - p_{RI}R_I^d$, and the value added of the natural resources provider is $p_{RI}R_I^d$. Summing the individual components, we obtain Y_I . This simple formula for value added reflects the fact that intermediate goods firms and natural resource providers do not purchase intermediate goods for use in their own production. Nominal GDP also includes primary production, which is the provision of natural resources to sectors other than manufacturing, and so we obtain the following equation for nominal GDP:

(26) $gdp = Y_I + Y_c + p_{RI}^t(R_I^s - R_I^d) + p_{Rc}^t(R_c^s - R_c^d).$

The uniform wage level (included a mark-up in the amount of k for profits) is determined by the equation

(27) $(1 + k)Lw = gdp.$

The ratio at which consumer goods and investment goods are demanded and traded in the market remains unspecified. Because we consider open economies, consumer preferences do not fix this ratio, as in Sachs and Warner (1999). This ratio is affected by two factors. First, the relative levels of technology (efficiency) in consumer goods production and investment goods production affect this ratio. If A_t is high relative to B_t, firms have a comparative advantage in investment goods production and will specialize in investment goods rather than in consumer goods. Second, because consumers favour local goods, the entry conditions for consumer goods producers are relatively strict in small countries, which decreases the share of consumer goods production in total production. The product ratio is determined by the following equation:

(28) $\frac{Y_c}{Y_c+Y_I} = \left(b_1 + b_2\theta^\eta \left(\frac{E(B_t)}{E(A_t)+E(B_t)}\right)^{1-\eta}\right)^m,$

where $0 < \eta < 1$ and $0 < m < 1$.

Setting parameters b1, b2, η and m in (26) to the appropriate levels in model simulations, the range within which the consumer goods ratio varies can be fixed to correspond to the typical range in our country data. According to formula (26), the country with a relatively high level of A_t also has a relatively large share of investment goods production.

Appendix A2.2 Model simulations

In this chapter, we present the results of simulations that use the theoretical model presented in chapter 2.1. Our aim is to consider the impacts of parameters that are thought to characterize the role of natural resources and geography in economic growth. We consider the change in each parameter separately.

We simulate the model using given values for parameters and initial values for the state of technology (A_0, $B_0 > 0$). We solve the static equilibrium of the model for the first period and consider the solved values, A_1 and $B_1 (> 0)$, as initial values for the next period. When moving from one period to the next, the values of other endogenous variables from the previous period are used as initial guesses for the current time simulation. This procedure is repeated t times. The static equilibrium of the model is solved using the non-linear system of 11 equations that are obtained from the model in chapter 2.1. Each equation corresponds to one endogenous[31] variable of the model.[32]

We bound production structure within a reasonable interval, as follows. When there are no differences in initial values and parameters among the sectors – which include the production of investment and intermediate goods as well as the consumer goods sector – we set parameter values b_1,b_2 in equation (26) so that $\frac{Y_c}{Y_I+Y_c} = 0.5$. Thus, for example, when m = 0.7, $\eta = 0$ and $A_t = B_t$, we set b_1 = 0.07 and $b_2 \approx 0.6$, which gives us the upper and lower bounds of the production structure. Specifically,

$$\lim_{t\to\infty}: \frac{Y_c}{Y_I+Y_c} = (b_1+b_2)^m \approx 0.76$$

when $\Delta A_t < \Delta B_t$, and

$$\lim_{t\to\infty}: \frac{Y_c}{Y_I+Y_c} = (b_1)^m \approx 0.16$$

[31] The endogenous variables are $c_I, c_c, w, Y_I, Y_c, x_I, x_c, A, B, L_I$ and L_c.
[32] This is accomplished with the Broyden method. This method is similar to the Newton-Raphson method, which uses a Jacobian to update initial guesses and approximate values for variables until a satisfactory tolerance is achieved. In the Broyden method, the Jacobian is updated, and it is no longer necessary to recalculate it after every iteration.

when $\Delta A_t > \Delta B_t$. However, in our model, it is also possible that the growth rate of the sectoral technology levels changes from $\Delta A_t < \Delta B_t$ to $\Delta A_t > \Delta B_t$ as t grows and that the production structure does not converge to either lower bound or upper bound given the assumed values for the parameter set.

In the reported simulations, we initially set the parameter values as follows:

α = 0.4, r = -2.3 (i.e., the elasticity of substitution = 3), σ = 0.7, λ = 0.8, m = 0.7, h_I = 0.6, h_c= 0.6, θ = 1, d_I = 1, d_c = 1, p_{RI}^t = 1, p_{Rc}^t = 1, p_{RI}^l = 1, p_{Rc}^l = 1, φ = 1.5, s = 0.5, R_I^s = 10, R_c^s = 10, k = 0.4, A_0 = 1, B_0 = 1, L = 100, b_1= 0.07, b_2= 0.6, η = 0, K_I=0, and t = 100.

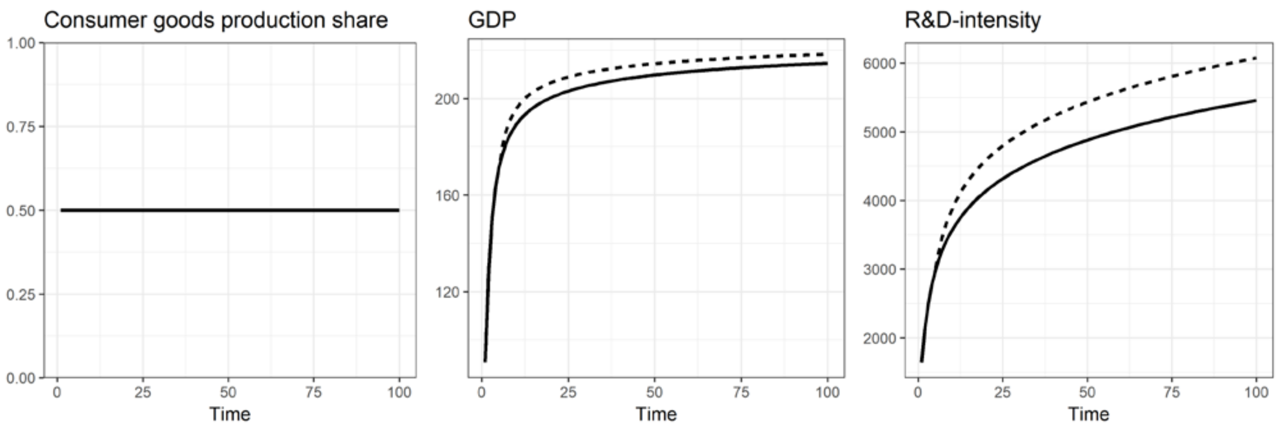
In each simulation, we investigate the impact of a change in the value of a parameter that is assumed to represent a productivity or growth impact of geography or natural resources. Our model is basically a Schumpeterian growth model. Thus, the basic scenario to which the alternative scenarios are compared is upward sloping in terms of GDP.

Changes in the efficiency parameter λ

We think that internal density strengthens agglomeration externalities, which in turn increases the value of efficiency parameter λ. In Figure 1, we depict the model behaviour when λ increases from 0.8 to 1.0.

Note that because λ is assumed to be the same for both sectors, output does not start to concentrate in either the consumer goods sector or the investment and intermediate goods sectors. The dotted (blue) line shows that GDP grows faster when λ is higher than the baseline level (solid line in Figure 1). R&D intensity across the economy also rises, as is expected.

Figure 1. The impact of an increase in λ. The solid line represents the case in which $\lambda = 0.8$, and the dotted represents the case in which $\lambda = 1$.

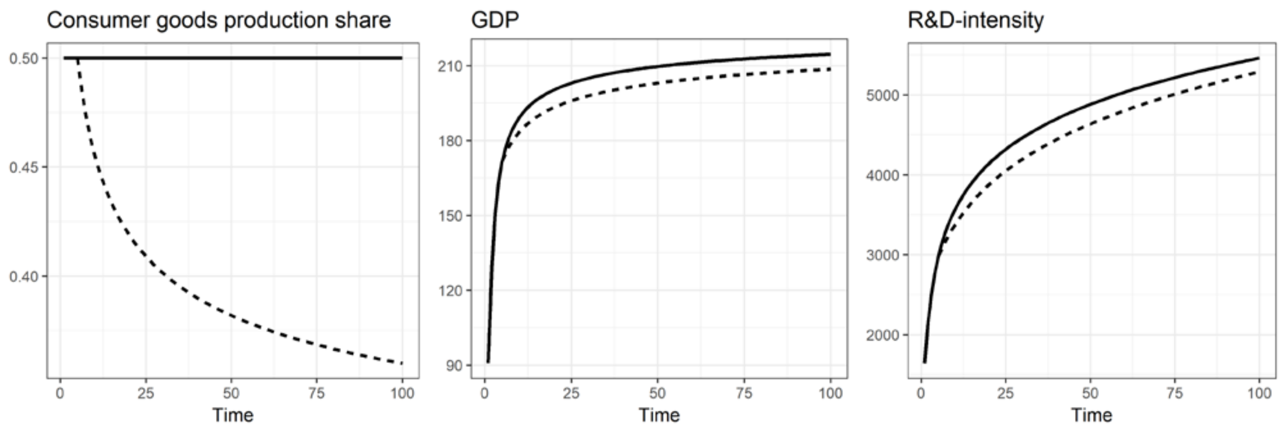


Changes in R&D cost parameters d_I and d_c

Next, we assume that d_I < d_c. This kind of asymmetry is thought to describe the fact that quality matters more in purchases of investment and intermediate goods than in purchases of consumer goods. Suppose that it is equally costly to improve the quality of investment and intermediate goods and consumer goods. Because quality matters more in the former goods group, it is cheaper to gain such quality improvements through R&D investments, which equally promote the sale of investment and intermediate goods and the sale of consumer goods.

The solid (red) line in Figure 2 represents the case in which d_c = d_I = 1. The dotted (blue) line is the case in which $d_c = 3 > d_I = 1$. Due to the assumed change, the economy starts to move away from the consumer goods sector, as shown in the simulation on the right-hand side of Figure 2. In fact, the sectoral structure of technologies $\frac{B_t}{B_t+A_t}$ decreases in the same way as $\frac{Y_c}{Y_c+Y_I}$ in Figure 2. That an increase in d_c would lower GDP or productivity and overall R&D intensity (in Figure 2) is rather self-evident.

Figure 2. The impact of an increase in d_c. The solid line represents the case in which $d_c = d_I = 1$, and the dotted line represents the case in which $d_c = 3$ and $d_I = 1$.

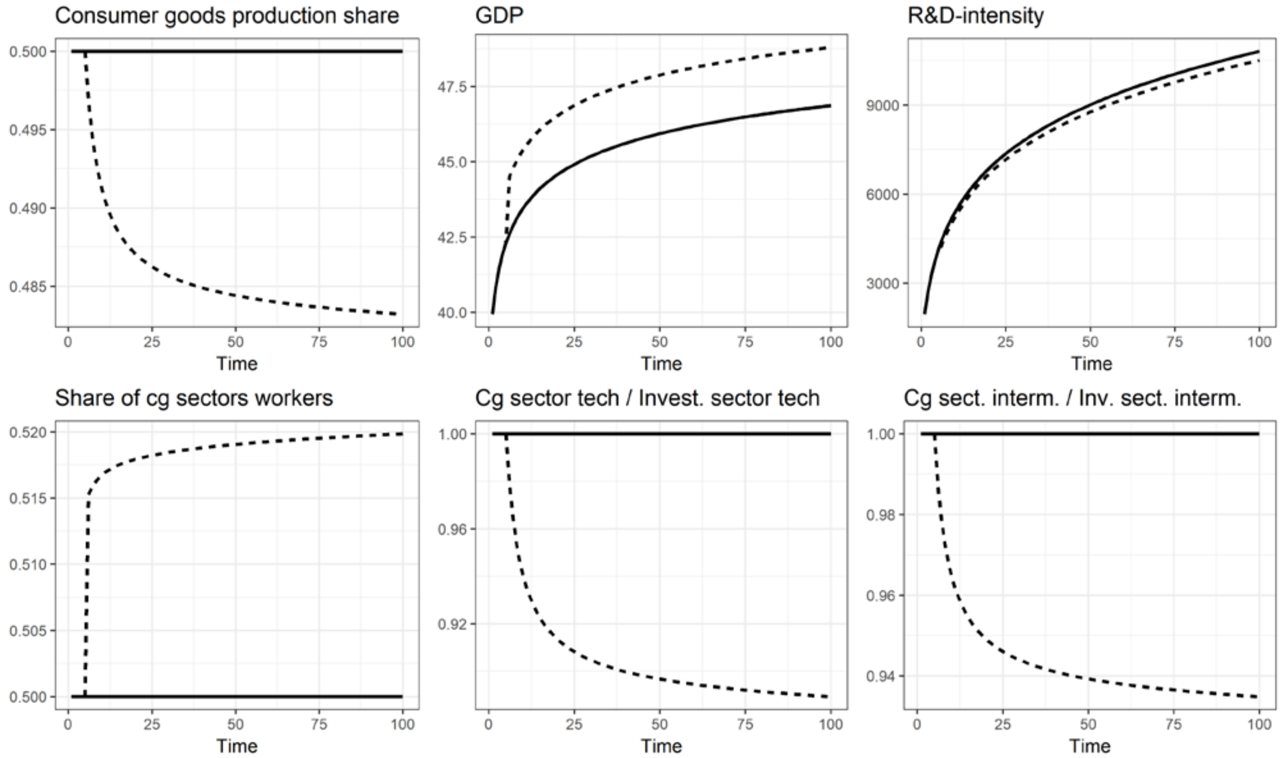


Changes in sectoral natural resources R_I^s and R_c^s

Next, we analyse the impacts of changes in the abundance of natural resources. In Figure 3, we depict the case in which the supply of investment goods sector resources R_I^s increase (in period 5).

We also assume that d_c = d_I and that $p_{RI}^t = p_{Rc}^t = 1.9 > p_{RI}^l = p_{Rc}^l = 1$. Note that it is crucial that global prices are higher than local prices. Then, the prices of natural resources for consumer goods production depart from the prices of natural resources for investment and intermediate goods production. Only then would an increase in natural resources have an effect on the sectoral shares of production and R&D activity.

Figure 3. The impact of an increase in R_I^s. The solid line represents the case in which d_c = d_I and $R_c^s = R_i^s$, and the dotted line represents the case in which d_c = d_I and $10 = R_c^s < R_i^s = 11$.



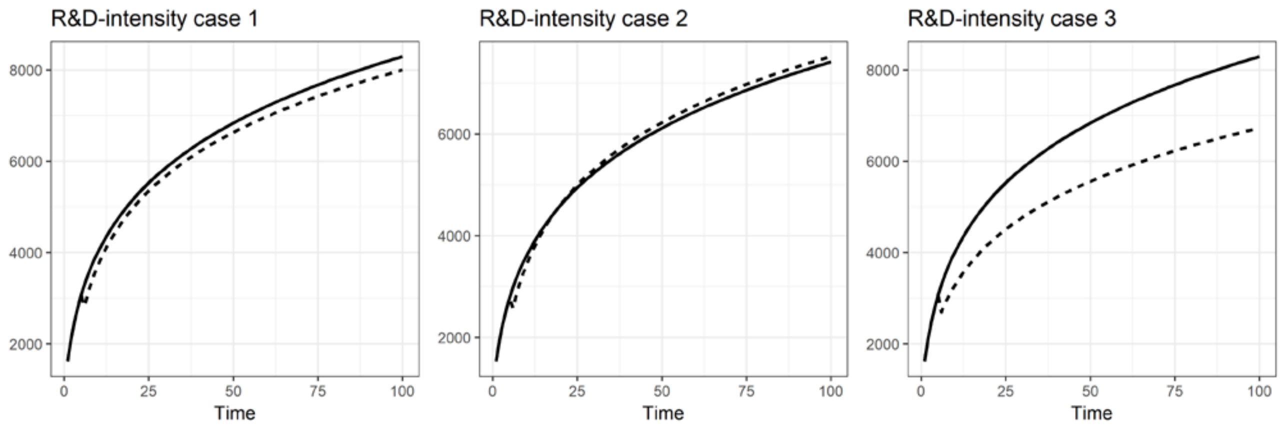
An increase in R_I^s causes the economy to concentrate more on investment goods production. More specifically, if marginal costs c_I drop below costs c_c, the production of intermediate goods x_I increases instantly and continues to grow faster than the production of intermediate goods x_c. An increase in R_I^s instantly reduces the labour (L_I) used in the production of investment goods because in the model, an increase in investment resources replaces labour input in the production of investment goods. An increase in investment resources also instantly increases GDP level; it also accelerates GDP growth. Note that the model is symmetric in the sense that if we had $d_c = d_I$ and $11 = R_c^s > R_i^s = 10$, the outcome of the model would be an exact mirror image of the graphs in Figure 3, except for the GDP curve, which would remain the same.

With an increase in R_I^s, innovative activity concentrates more on the investment and intermediate goods sector. Nonetheless, an increase in R_I^s replaces innovative inputs with natural resources in the production of investment and intermediate goods, which lowers the average R&D intensity of the economy.

R&D intensity and natural resources

Next, we consider more closely how an increase in natural resources impacts R&D intensity. It was previously shown in Figure 3 that an increased supply of natural resources tends to lower R&D intensity. On the left-hand side in Figure 4, we consider the case in which $d_c = 3$ > $d_I = 1$ and R_i^s increases from 10 to 15 after period 5. The simulated adjustment process is similar to the process described in Figure 3. Although d_I is now much lower than d_c, an increase in R_i^s does not encourage a sufficient increase in the R&D inputs used in the production of investment and intermediate goods to cause overall R&D intensity to rise. In contrast, it falls. In the centre-most part in Figure 4, $d_c = 7 > d_i = 1$, meaning that R&D costs are initially very low in investment and intermediate goods production compared with R&D costs in consumer goods production. Therefore, an assumed increase in R_i^s, which shifts resources from consumer goods production to investment and intermediate goods production, raises the average R&D intensity. This result requires that the elasticity of substitution between labour and intermediate good input is sufficiently small. If this elasticity increases from 0.3 to 1 (which is inherent in the Cobb-Douglas production function), an increase in R_i^s induces a firm to substitute R&D input for natural resource input to the extent that R&D intensity decreases.

Figure 4. Left: $d_c = 3 > d_i = 1$ and an increase in R_i^s from 10 to 15 after period 5. Middle: $d_c = 7 > d_I = 1$ and an increase in R_I^s from 10 to 15 after period 5. Right: $d_c = 3 > d_I = 1$ and an increase in R_c^s from 10 to 15 after period 5.

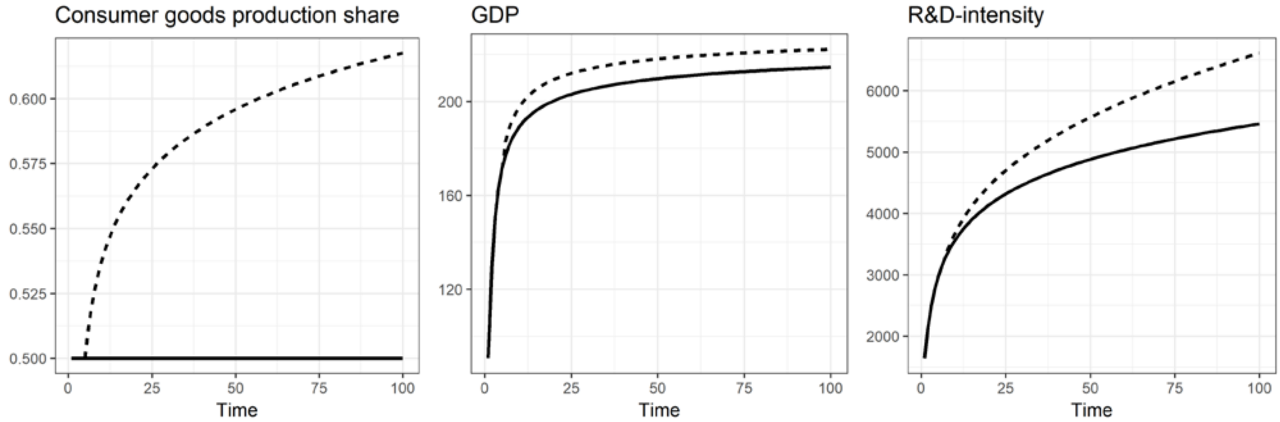


On the right-hand side in Figure 4, $d_c = 3 > d_i = 1$ and R_c^s is assumed to increase from 10 to 15 after period 5. In this case, both the difference between d_c and d_I and the increase in R_c^s significantly reduce average R&D intensity because the activity expands in the sector with high R&D costs.

Changes in the home market size parameter θ.

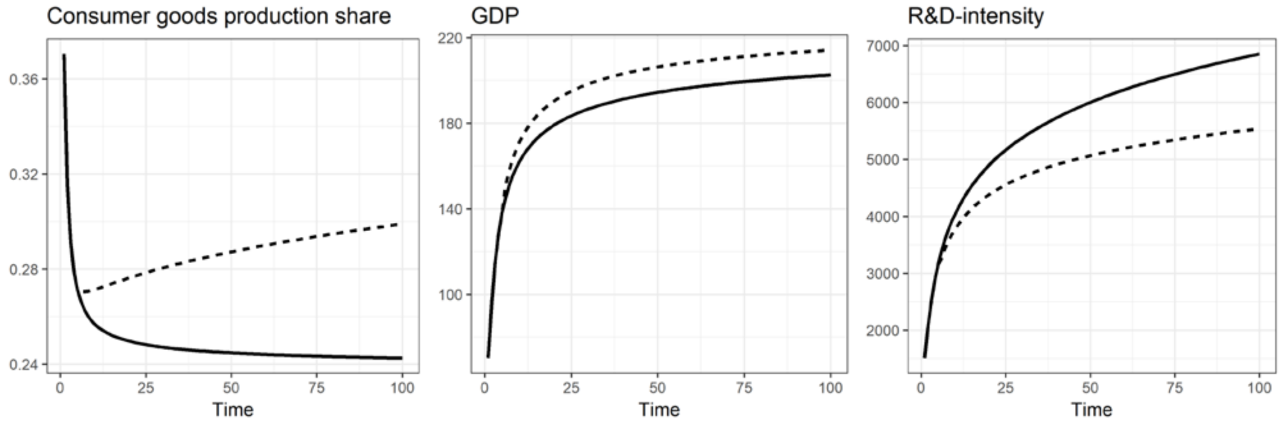
Next, we examine how a change in θ affects the model. This parameter captures the impact of home market size. Figure 5 depicts the case in which parameter θ increases from 1 to 1.05 in period 5. Other sectoral parameters remain equal. For simplicity, we have assumed that $K_I = K_c = 0$.

Figure 5. The impact of a change in θ. For the solid line, $\theta = 1$, and for the dotted line, $\theta = 1.05$.



When θ grows, the profitability of the production of intermediate product x_c increases, and the incentive to invest in R&D technology B_t becomes stronger. Thus, a jump in θ expands GDP and raises R&D intensity. We also investigate the case in which d_I = 1 and d_c = 3. Under these assumptions, an increase in θ still increases GDP and accumulates consumer goods production at the expense of investment goods production. Now, R&D intensity also drops due to the relatively high R&D costs in the consumer goods production. This is depicted n Figure 6.

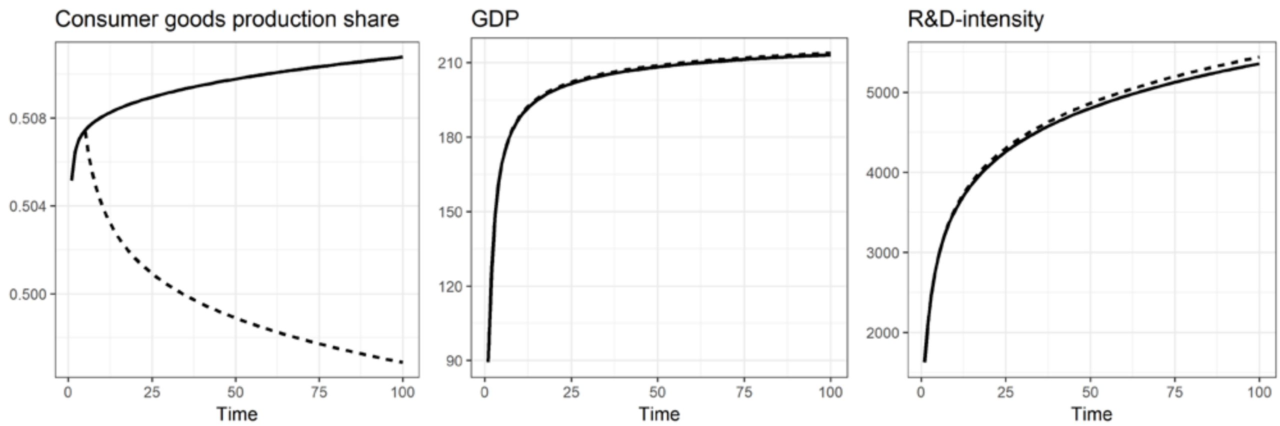
Figure 6. The impact of a change in θ when d_I = 1 and d_c = 3. For the solid line, $\theta = 1$, and for the dotted line, $\theta = 1.05$ after period 5.



Increase in fixed costs $K_I > 0$ in investment sector and increasing price competition $\theta < 1$ in consumer goods sector.

Consider now a different interpretation for θ. The profits of the intermediate goods monopolist from the production of investment goods and consumer goods are $\pi_{xI} = p_{xI}x_I - c_I x_I - K_I$ and $\pi_{xc} = \theta p_{xc}x_c - c_c x_c$, respectively. Now there are fixed costs in the investment sector, $K_I > 0$, but not in the consumer goods sector. There are, therefore, increasing returns to scale in the investment sector, whereas in the consumer goods sector, constant returns remain. Additionally, we assume that in the consumer goods sector, the demand for intermediate products sets prices as the marginal product $\frac{\partial Y_c}{\partial x_c}$. However, the price is lowered by the producers' anticipated increase in price competition $\theta < 1$, which they account for when maximizing profits. This case is depicted in Figure 7 with $\theta = 0.997$, $K_I = 0.8$, and after period 5, the fixed costs drop to $K_I = 0.1$.

Figure 7. Drop in fixed costs $\mathrm{K_I} > 0$ when $\theta < 1$



The solid line represents an economy with $\theta = 0.997$, $K_I = 0.8$. First, it starts to focus on consumer goods production since the fixed effects are so high that it is optimal to focus on consumer goods even though $\theta = 0.997$. The dotted line represents an economy where the fixed costs drop to $K_I = 0.1$. As a consequence, the economy starts to focus on the investment goods sector. There therefore exists a relationship between the parameters θ and K_I that determines whether increasing returns to scale dominate price competitiveness.

A2.3. A model for education and geography

It is widely believed that education enhances both human capital and economic growth.[33] Education is also seen as a device used to signal individual skill, which is characterized by information asymmetry (Spence, 1973). In this study, we rely on the human capital theory. The support of a public authority for schooling is thus explained by the motive to internalize positive externalities related to education or by the motive to promote equity. The former case, on which we focus, is characterized by a sort of market failure. Positive externalities are thus related to the aggregation of human capital. The spillover effects of human skills may cause the social returns from local education to be higher than the private returns. Without public support, education would remain below its optimal level. Next, we investigate the impacts of amenities that affect the immigration and emigration of highly skilled people and thus influence the social returns to education

[33] See, e.g., Glaeser et al. (2004).

and ultimately impact public expenditure on education.[34] In our analysis, the public authority has two instruments: school expenditure and investment in amenities to attract highly skilled people.[35] Geography impacts the intensity of these amenities. For example, the agglomeration related to geographic density can be assumed to increase R&D investments and the demand for highly skilled people, which increases the efficiency of investments in amenities to attract highly skilled people.

Let w be the extra wage paid to highly educated individuals. Assume that a highly educated individual becomes a top executor with probability p and that a top executor is paid (1+k)w, k > 0. Let c be public expenditure on higher education per student. The private cost to educate oneself beyond the basic level is h(c), with h'(c) < 0 and h''(c) > 0. Individuals are heterogeneous, and their abilities are uniformly distributed in the range (0,1). Heterogeneity is assumed to be reflected in additional education costs. For the cleverest person, this cost is zero. The expected value of private returns from schooling over individual characteristics is

(1) $E(PR) = (1-p)w + p(1+k)w - h(c) - d\int_0^1 v dv,$

where d is the cost parameter related to extra education costs. An individual knows her own talent when the school decision is made. But she does not know her success later in the working life. An individual of type v, thus, decides to participate in higher education, if

$PR = (1-p)w + p(1+k)w - h(c) - dv \geq 0.$

From the above condition, one obtains a cut-off point

$$v^* = \frac{(1-p)w + p(1+k)w - h(c)}{d}.$$

When v < v*, an individual goes to higher education. From the above condition, it follows that

$\frac{\partial v^*}{\partial c} = -\frac{h'(c)}{d} > 0$ and $\frac{\partial^2 v^*}{\partial c^2} = -\frac{h''(c)}{d} < 0$.

In other words, the share of those who go to higher education increases with c at decreasing rate.

This fraction, denoted by s (instead of v^*), can then be defined as the function

[34] Roback (1988) and Moretti (2004, 2006) consider the regional impacts of exogenous amenities that attract highly educated people. In their analyses, regional differences in the amenities that affect labour supply and the technology shocks that impact the demand for highly skilled people may cause regions to differ from each other with respect to wages and productivity.
[35] Various policies to induce highly educated people to immigrate are analysed in Bertoli et al., 2011.

(2) s = s(c,#),

where # captures the impact of parameters other than c, and s'(c) > 0 and s''(c) < 0. Suppose that the social benefits of high education are the same as wages.

For simplicity, we assume that only top executors immigrate and emigrate. The immigration and emigration of top executors are affected by investments e in amenities to attract these people. We suppose that a fraction ψm(e) of all people who obtained higher education and became top executors emigrate. Then, m'(e) < 0 and m''(e) > 0. Accordingly, θn(e) is the number of top-level talents that immigrate, so that n'(e) > 0 and n''(e) < 0. Above ψ and θ are country-specific sensitivity parameters.

Let N be the total population, which comprises people with only basic education plus people with higher education. The social returns to schooling are then

(3) SR = Ns(c,#)[(1-p)w + p(1+k)w+(1-ψm(e))p(1+k)w – c] + θn(e)(1+k)w - e.

We also assume that public investments are restricted by a constraint B so that

(4) B = Ns(c,#)c + e.

We suppose that it pays to invest in schooling so that $\frac{\partial SR}{\partial c} > 0$ at c = 0 with any e ≥ 0 and with given k,p and w. From this it follows that at the possible extremal that maximizes (1), (1-p)w + p(1+k)w - c +(1-ψm(e))p(1+k)w > 0.

The public planner thus maximizes SR in (3) with respect to c and e subject to constraint (4). The first-order conditions are obtained by maximizing the Lagrangian function with respect to c and e, as follows:

(5) L = Ns(c,#)[(1-p)w + p(1+k)w - c +(1-ψm(e))p(1+k)w]+ θn(e)(1+k)w - e + λ(B – Ns(c,#)c - e)

Differentiating Lagrangian in (5) with respect c and e one obtains the first order conditions

(i) N[s'(c,#)[(1+pk)w – c+(1-ψm(e))p(1+k)w] – Ns(c,#) - λN[s'(c,#)c + s(c,#)] = 0;

(ii) - Ns(c,#)ψm'(e)p(1+k)w+θn'(e)(1+k)w -1 - λ = 0; and

(iii) B = Ns(c,#)c + e.

To consider the validity of the second order conditions, we derive from (i) and (ii) the expressions

(6) $\frac{\partial^2 L}{\partial c^2} = Ns''(c,\#)[(1-p)w + p(1+k)w - c + (1-\psi m(e))p(1+k)w] - Ns'(c,\#) - Ns'(c,\#) - \lambda N[s'(c,\#) + s''(c,\#)c + s'(c,\#)]$.

(7) $\frac{\partial^2 L}{\partial e^2} = -Ns(c,\#)\psi m''(e)p(1+k)w + \theta n''(e)(1+k)e < 0$

(8) $\frac{\partial^2 L}{\partial c \partial e} = -Ns'(c,\#)\psi m'(e)p(1+k)w > 0$

To examine more closely second order partial derivative with respect to c, we solve using (6) and (i), the following equation

$$\frac{\partial^2 L}{\partial c^2} = \{Ns(c,\#)]s''(c,\#)\big[w(1+pk) - c + \big(1 - \psi\mathrm{m(e)}\big)\mathrm{p(1+k)w}\big] - 2Ns'(c,\#)[s'(c,\#)c + s(c,\#)] - 2N\big(s'(c,\#)\big)^2 - N\big(s'(c,\#)\big)^2[w(1+pk) - c + \big(1 - \psi\mathrm{m(e)}\big)\mathrm{p(1+k)w}] + \mathrm{Ns(c,\#)s''(c,\#)]}\}\frac{1}{[s'(c,\#)c + s(c,\#)]}$$

of which left-hand side is clearly negative.

Notice also that

$$\frac{\partial B}{\partial c} = Ns'(c,\#)c + Ns(c,\#) > 0 \text{ and } \frac{\partial B}{\partial e} = 1 > 0,$$

Now we can show that the determined of the bordered Hessian

$$\det H_3 = \begin{vmatrix} \frac{\partial^2 L}{\partial c^2} & \frac{\partial^2 L}{\partial c \partial e} & \frac{\partial B}{\partial c} \\ \frac{\partial^2 L}{\partial c \partial e} & \frac{\partial^2 L}{\partial e^2} & \frac{\partial B}{\partial e} \\ \frac{\partial B}{\partial c} & \frac{\partial B}{\partial e} & 0 \end{vmatrix},$$

which can be written as,

$$detH_3 = \frac{\partial^2 L}{\partial c^2}\left(\frac{\partial B}{\partial e}\right)^2 + \frac{\partial^2 L}{\partial c \partial e}\left(\frac{\partial B}{\partial e}\right)\left(\frac{\partial B}{\partial c}\right) + \frac{\partial B}{\partial c}\left(\frac{\partial^2 L}{\partial c \partial e}\frac{\partial B}{\partial e} - \frac{\partial B}{\partial c}\frac{\partial^2 L}{\partial e^2}\right),$$

is positive so that the second order conditions for maximization are valid.

From (i) – (iii), one also obtains the result

(9) N[s'(c,#)((1+pk)w − c +(1-ψm(e))p(1+k)w] − Ns(c,#) + N[s'(c,#)c + s(c,#)][Ns(c,#)ψm'(e)p(1+k)w - θn'(e)(1+k)w + 1] = 0,

where B = Ns(c,#)c + e ↔ e = B - Ns(c,#)c. Let A denote the left-hand side of (9). The validity of the second order conditions guarantees that $\frac{\partial A}{\partial c} < 0$.

From (9), one also obtains that

(10) $\frac{\partial A}{\partial \psi} = -Ns'(c,\#)m(e)p(1+k)w + N[s'(c,\#)c + s(c,\#)Ns(c,\#)m'(e)p(1+k)w < 0,$

so that $\frac{\partial c}{\partial \psi} = -\frac{\frac{\partial A}{\partial \psi}}{\frac{\partial A}{\partial c}} < 0$, indicating that an increase in the outflow of high skills decreases school expenditure.

Similarly, we obtain

(11) $\frac{\partial A}{\partial \theta} = -N[s'(c,\#)c + s(c,\#)][(1+k)wn'(e) < 0$.

This lets us to conclude that $\frac{\partial c}{\partial \theta} = -\frac{\frac{\partial A}{\partial \theta}}{\frac{\partial A}{\partial c}} < 0$, which tells us that vitalized inflow of highly skilled people decreases schooling expenditures.

Together, the results tell us that an increase in the global movement of talent decreases school expenditure. Considering the period 1995 – 2011, during which the importance of R&D investments was emphasized, it is sensible to assume that the countries with high density (agglomeration) started to attract talents for the increased intensity of R&D, which had a greater impact than the deceleration of the outflow of talent because this outflow was already slight. Consequently, parameter θ would increase but ψ would be unchanged. Accordingly, c would fall. If the sparsely populated countries in the sample could not attract talents but did not suffer significant “brain drain”, then the incentives to invest in public education remain strong. In all, it possible that in sparsely populated and/or remote countries, the turnover of talent through emigration and immigration is lower compared with that in densely populated and/or centrally located countries. According to this, the incentives to make public investments in education are stronger in the former group.

Table A1 Summary statistics of the data *

Isocode	Prod	Rank	Prod 90	Rank	Share of c-goods	Rank	Milit	Rank	R&D-intensity	Rank	Population	Rank	Agri	Rank	Dens1	Rank	Dens2	Rank	Extdens1	Rank
ARG	22.3	30	9.1	38	33.5%	2	1.0%	36	0.2%	39	33.6	16	46.7%	22	12	37	54	29	34	38
AUS	42.5	16	30.8	12	25.4%	29	1.9%	14	1.5%	17	17.5	19	60.5%	9	2	42	17	41	24	41
AUT	45.2	12	29.8	14	23.6%	37	0.9%	39	2.1%	10	7.8	29	42.2%	26	95	21	72	25	163	11
BEL	55.5	2	40.7	2	27.0%	21	1.2%	32	1.8%	14	10.1	25	45.3%	24	332	4	147	14	343	1
BGR	14.3	38	14.1	32	29.8%	10	2.3%	10	0.2%	41	8.6	28	55.6%	12	78	24	68	26	117	21
BRA	11.8	40	6.7	39	30.3%	9	1.6%	20	0.4%	29	155.4	3	29.5%	32	19	33	142	16	33	39
CAN	43.9	15	35.2	8	24.8%	30	1.2%	33	1.6%	16	28.3	17	7.5%	41	3	40	24	36	187	8
CHE	47.2	9	38.4	4	23.4%	39	1.0%	38	2.6%	7	6.8	30	40.4%	29	172	9	90	22	202	4
CHL	16.7	34	10.2	36	27.7%	18	2.5%	8	0.2%	38	13.6	21	21.0%	35	18	34	41	30	33	40
CHN	4.4	42	1.9	42	24.0%	32	1.9%	13	0.3%	32	1186.9	1	54.8%	15	126	11	1000	1	55	33
CZE	27.0	24	22.2	22	23.9%	34	1.6%	19	0.6%	25	10.3	23	55.4%	14	134	10	98	18	180	9
DEU	51.4	6	32.6	11	23.7%	36	1.4%	29	2.1%	9	80.1	7	49.6%	21	229	6	358	4	194	5
DNK	49.8	7	35.7	7	28.9%	13	1.5%	26	2.6%	6	5.2	32	64.8%	5	122	13	66	27	160	12
ESP	39.8	17	28.3	18	28.1%	17	1.5%	24	0.9%	23	39.4	14	60.6%	8	79	23	146	15	109	25
EST	18.5	33	11.3	34	27.3%	20	1.6%	22	0.5%	26	1.5	41	30.8%	31	36	30	19	39	128	16
FIN	44.0	14	28.8	17	20.6%	42	1.3%	31	3.4%	3	5.0	33	7.9%	40	17	35	23	37	110	24
FRA	52.0	5	37.4	5	26.7%	23	2.5%	7	2.1%	11	59.0	8	55.5%	13	108	18	212	9	203	3
GBR	45.4	11	30.2	13	26.1%	25	2.4%	9	1.9%	12	57.4	10	73.9%	1	237	5	310	5	191	7
GRC	29.6	23	19.7	26	32.3%	5	3.0%	5	0.4%	28	10.3	24	71.2%	2	80	22	76	24	97	27
HUN	22.4	29	13.4	33	24.7%	31	1.3%	30	0.5%	27	10.4	22	69.7%	3	115	14	90	21	149	14
IRL	46.6	10	24.6	21	25.4%	28	0.7%	40	1.3%	18	3.6	37	67.6%	4	52	27	36	33	156	13
ISL	36.1	19	29.0	16	33.1%	3	0.3%	42	3.5%	2	0.3	42	19.0%	36	3	41	2	42	85	29
ISR	35.2	20	29.5	15	27.6%	19	7.5%	1	3.3%	4	4.8	34	26.7%	33	223	7	92	20	72	31
ITA	45.0	13	34.2	9	26.8%	22	1.8%	16	0.9%	22	57.1	11	54.7%	16	194	8	275	6	113	22
JPN	36.4	18	25.5	20	23.9%	33	1.0%	37	3.2%	5	123.2	5	15.4%	37	338	3	537	2	35	37
KOR	23.1	28	10.7	35	22.1%	40	2.5%	6	1.9%	13	43.7	13	22.2%	34	453	1	369	3	71	32
LTU	20.9	31	14.9	30	32.4%	4	1.0%	35	0.3%	31	3.7	36	53.0%	18	59	26	38	31	123	19
LVA	16.2	35	15.1	29	29.4%	11	1.2%	34	0.3%	35	2.6	39	41.1%	28	42	29	27	35	123	18
MEX	15.4	36	14.2	31	28.2%	15	0.5%	41	0.2%	37	89.1	6	54.5%	17	46	28	169	12	50	34
NLD	53.6	4	39.2	3	25.7%	27	1.5%	25	1.7%	15	15.1	20	58.9%	11	448	2	197	10	316	2
NOR	62.4	1	36.4	6	26.2%	24	1.8%	18	1.2%	19	4.3	35	2.8%	42	12	38	20	38	128	17
NZL	32.8	22	26.4	19	34.2%	1	1.4%	27	1.0%	21	3.5	38	60.2%	10	13	36	18	40	24	42
POL	19.5	32	9.5	37	28.2%	16	1.9%	15	0.3%	34	38.4	15	61.2%	7	125	12	182	11	133	15
PRT	26.1	25	15.8	28	30.7%	7	2.0%	12	0.7%	24	10.0	26	43.0%	25	109	17	87	23	104	26
ROU	13.0	39	6.2	40	32.2%	6	2.1%	11	0.2%	40	23.3	18	64.5%	6	102	19	125	17	111	23
RUS	15.0	37	21.5	23	28.3%	14	3.7%	2	0.3%	33	148.2	4	13.4%	38	9	39	96	19	82	30
SWE	47.7	8	33.3	10	21.9%	41	1.6%	21	3.8%	1	8.7	27	8.2%	39	21	32	34	34	121	20
SVK	26.1	26	19.8	25	23.5%	38	1.8%	17	0.3%	30	5.3	31	50.4%	20	111	16	64	28	191	6
SVN	33.5	21	20.5	24	23.9%	35	1.4%	28	1.1%	20	2.0	40	33.3%	30	99	20	37	32	171	10
THA	7.6	41	4.1	41	29.3%	12	1.5%	23	0.1%	42	57.7	9	41.7%	27	113	15	214	8	41	36
TUR	23.2	27	18.3	27	30.5%	8	3.2%	4	0.3%	36	55.8	12	51.9%	19	73	25	168	13	92	28
USA	54.8	3	41.6	1	25.8%	26	3.7%	3	2.3%	8	258.0	2	46.4%	23	28	31	221	7	46	35
Mean	**32.7**	-	**23.0**	-	**27.2%**	-	**1.8%**	-	**1.3%**	-	**64.4**	-	**43.6%**	-	**112**	-	**144**	-	**121**	-
S.t. Dev.	**15.3**	-	**11.2**	-	**3.4%**	-	**1.2%**	-	**1.1%**	-	**184.9**	-	**19.6%**	-	**113**	-	**177**	-	**72**	-
Min	**4.4**	-	**1.9**	-	**20.6%**	-	**0.3%**	-	**0.1%**	-	**0.3**	-	**2.8%**	-	**2**	-	**2**	-	**24**	-
Max	**62.4**	-	**41.6**	-	**34.2%**	-	**7.5%**	-	**3.8%**	-	**1186.9**	-	**73.9%**	-	**453**	-	**1000**	-	**343**	-
Median	**33.2**	-	**23.4**	-	**0.26892**	-	**0.0**	-	**0.0**	-	**12.0**	-	**0.5**	-	**87.4**	-	**90.1**	-	**115.1**	-

*Table A1 continued

Isocode	Extdens2	Rank	Extdens3	Rank	Education	Rank	Forest	Rank	Gas&Oil	Rank	Minerals	Rank	Univ	Rank	GDP/Population	Rank	GDP/Population 90	Rank
ARG	53	38	0.6	36	177	38	12.5 %	37	722	9	1619	18	0.1	37	15114	31	5945	39
AUS	51	39	0.4	39	238	26	16.7 %	34	941	7	73142	1	4.8	11	37853	8	26768	7
AUT	164	5	1.3	13	253	14	45.9 %	10	71	21	412	28	3.3	16	37654	9	25241	16
BEL	195	2	1.9	5	243	21	22.0 %	30	0	40	15	40	4.2	14	35624	13	26031	11
BGR	151	11	1.0	24	214	32	30.2 %	21	19	29	1801	16	0.1	39	11050	38	12026	32
BRA	48	41	0.6	34	145	42	64.8 %	5	121	18	4044	10	0.2	35	9930	40	5890	40
CAN	102	33	3.5	1	255	13	38.3 %	12	2567	4	12362	4	5.2	8	38050	7	30558	4
CHE	153	10	1.2	17	275	3	29.3 %	22	0	39	66	37	11.9	2	44583	3	37685	1
CHL	49	40	0.4	38	189	35	20.7 %	31	252	14	36758	3	0.5	29	13629	35	7591	37
CHN	85	35	1.0	22	174	39	17.2 %	33	136	17	1774	17	0.1	38	5913	42	2386	42
CZE	156	9	1.4	11	262	10	34.0 %	16	7	31	499	26	0.8	28	23798	25	22117	18
DEU	148	13	2.2	3	258	12	32.4 %	19	87	20	717	24	2.6	19	35684	12	25732	12
DNK	142	18	1.0	23	277	2	13.0 %	35	1282	6	174	32	5.7	7	37143	11	26616	8
ESP	110	31	1.4	12	202	34	28.9 %	24	11	30	655	25	2.0	21	28915	20	17690	24
EST	142	19	0.8	31	268	7	52.2 %	7	0	38	1	42	1.2	26	16121	29	11420	33
FIN	134	24	1.2	18	261	11	72.2 %	1	0	34	10469	5	4.5	13	34148	15	25372	14
FRA	146	14	2.6	2	238	27	26.7 %	27	29	27	248	30	2.9	17	32911	18	25324	15
GBR	139	21	2.0	4	238	25	11.6 %	38	600	11	221	31	6.7	5	35557	14	24813	17
GRC	136	22	0.9	27	216	31	26.1 %	29	51	25	1916	14	1.3	25	26298	22	16715	26
HUN	156	8	1.2	19	241	23	20.3 %	32	256	13	57	38	0.9	27	18214	28	12902	30
IRL	122	28	1.1	21	222	30	7.2 %	40	176	15	2016	13	5.0	10	38333	6	18006	23
ISL	90	34	0.7	33	223	29	0.2 %	42	0	36	59451	2	8.2	4	38893	5	31286	3
ISR	143	17	0.4	41	245	20	6.3 %	41	118	19	3380	11	15.4	1	30584	19	21681	19
ITA	128	26	1.2	16	209	33	26.3 %	28	154	16	126	34	2.0	22	33260	17	25456	13
JPN	115	30	0.4	40	267	8	68.4 %	3	6	32	66	36	2.3	20	33832	16	27081	6
KOR	206	1	0.6	35	241	22	65.9 %	4	0	42	167	33	2.7	18	25741	23	12048	31
LTU	157	6	0.9	26	248	18	31.2 %	20	20	28	16	39	0.3	31	14874	32	13023	29
LVA	146	15	0.9	25	251	17	51.2 %	8	0	41	8	41	0.4	30	13835	33	16427	27
MEX	54	37	0.8	32	187	36	35.7 %	15	4199	3	2135	12	0.0	42	12648	37	10311	34
NLD	187	3	1.8	6	269	6	10.3 %	39	1581	5	447	27	5.2	9	40061	4	26334	10
NOR	124	27	1.5	8	284	1	33.2 %	17	24463	1	6278	6	4.6	12	46674	2	26539	9
NZL	44	42	0.3	42	252	15	37.0 %	14	718	10	4078	9	3.5	15	28440	21	21018	20
POL	145	16	1.5	9	247	19	29.1 %	23	61	23	1280	20	0.3	32	15399	30	8205	36
PRT	106	32	0.8	29	163	40	37.3 %	13	0	37	820	21	1.5	24	23301	26	14578	28
ROU	150	12	1.2	20	223	28	27.8 %	26	394	12	324	29	0.1	41	10600	39	6810	38
RUS	120	29	1.6	7	270	5	49.4 %	9	6316	2	1826	15	0.2	36	13643	34	19588	21
SWE	130	25	1.4	10	270	4	68.4 %	2	0	35	4909	7	6.6	6	37266	10	27951	5
SVK	170	4	1.3	14	251	16	40.0 %	11	45	26	811	22	0.3	33	18281	27	17072	25
SVN	157	7	0.8	30	240	24	59.4 %	6	0	33	774	23	1.6	23	24903	24	19142	22
THA	139	20	0.5	37	147	41	28.6 %	25	66	22	121	35	0.1	40	9416	41	5024	41
TUR	135	23	1.3	15	178	37	12.6 %	36	60	24	4420	8	0.2	34	12694	36	10125	35
USA	61	36	0.8	28	265	9	33.0 %	18	913	8	1592	19	8.2	3	47201	1	36398	2
Mean	**126**	-	**1.2**	-	**233**	-	**32.7 %**	-	**1106**	-	**5762**	-	**3.0**	-	**26383**	-	**19117**	-
S.t. Dev.	**41**	-	**0.6**	-	**37**	-	**18.4 %**	-	**3885**	-	**15031**	-	**3.4**	-	**11831**	-	**8903**	-
Min	**44**	-	**0.3**	-	**145**	-	**0.2 %**	-	**0**	-	**1**	-	**0.0**	-	**5913**	-	**2386**	-
Max	**206**	-	**3.5**	-	**284**	-	**72.2 %**	-	**24463**	-	**73142**	-	**15.4**	-	**47201**	-	**37685**	-
Median	**137.3**	-	**1.1**	-	**242.3**	-	**0.3**	-	**68.6**	-	**815.3**	-	**2.0**	-	**27369.1**	-	**19364.9**	-

Table A2 Correlation Matrix

	Prod	R&D-int	C-goods	Edu	Milit	Pop	Agri	Forest	Gas&Oil	Minerals	Univ	Dens1	Dens2	Extdens1	Extdens2	Extdens3	GDPpop	Prod90	GDPpop90
Prod	1.00																		
R&D-intensity	0.81	1.00																	
Share of c-goods	-0.31	-0.50	1.00																
Education	0.61	0.59	-0.43	1.00															
Milit	-0.09	-0.09	-0.03	0.01	1.00														
Population	-0.32	-0.21	-0.13	-0.36	0.35	1.00													
Agri	-0.18	-0.29	0.32	-0.26	0.03	0.14	1.00												
Forest	-0.12	-0.16	-0.36	0.11	0.30	0.32	-0.16	1.00											
Gas&Oil	-0.06	-0.26	0.23	-0.09	0.23	0.44	0.13	-0.07	1.00										
Minerals	0.08	0.10	0.03	-0.20	0.05	0.05	-0.31	-0.32	0.32	1.00									
Univ	0.84	0.92	-0.38	0.57	-0.01	-0.35	-0.24	-0.23	-0.22	0.09	1.00								
Dens1	0.01	0.05	-0.24	0.04	0.23	0.21	0.44	0.19	-0.16	-0.60	0.00	1.00							
Dens2	-0.22	-0.12	-0.23	-0.23	0.38	0.82	0.35	0.33	0.22	-0.31	-0.24	0.73	1.00						
Extdens1	0.19	0.17	-0.35	0.42	0.12	-0.24	0.02	0.10	-0.36	-0.55	0.16	0.63	0.19	1.00					
Extdens2	0.42	0.28	-0.32	0.46	-0.08	-0.30	0.06	-0.02	-0.24	-0.40	0.28	0.42	0.03	0.80	1.00				
Extdens3	0.32	0.18	-0.34	0.30	-0.03	0.13	-0.07	0.07	0.05	-0.14	0.12	0.16	0.18	0.51	0.81	1.00			
GDPpop	0.96	0.88	-0.37	0.62	-0.13	-0.32	-0.24	-0.16	-0.10	0.10	0.92	-0.02	-0.23	0.16	0.34	0.23	1.00		
Prod90	0.91	0.80	-0.32	0.61	-0.03	-0.17	-0.20	-0.14	0.02	0.09	0.82	-0.02	-0.13	0.16	0.41	0.40	0.90	1.00	
GDPpop90	0.85	0.83	-0.39	0.68	-0.12	-0.23	-0.28	-0.14	-0.11	0.08	0.83	-0.10	-0.22	0.15	0.34	0.30	0.90	0.94	1.00

Table A3, OLS estimates of the model in the 5th column in Tables 2-5.

	(1) Prod	(2) Prod	(3) R&D	(4) R&D	(5) Cg	(6) Cg	(7) Edu	(8) Edu
Time-varying endogenous variables								
R&D-intensity	0.250**	-	-	-	-0.042´	-	0.061*	-
	(0.047)				(0.024)		(0.029)	
Share of c-goods	-0.805**	-	-0.812	-	-	-	-0.544**	-
	(0.167)		(0.576)				(0.124)	
Education	0.023	-	1.023´	-	-0.359**	-	-	-
	(0.162)		(0.616)		(0.057)			
Time-invariant exogenous variables								
Agri	0.256**	0.068	-0.193	-0.279	0.035	0.10	-0.054*	-0.074
	(0.061)	(0.181)	(0.143)	(0.207)	(0.038)	(0.088)	(0.023)	(0.051)
Gas&Oil	0.002	-0.011	-0.067**	-0.065*	0.006´	0.004	0.017**	0.007
	(0.010)	(0.018)	(0.020)	(0.026)	(0.003)	(0.007)	(0.004)	(0.008)
Forest	-0.007	0.026	-0.103	-0.022	-0.022	-0.029	0.005	0.028
	(0.028)	(0.051)	(0.071)	(0.094)	(0.017)	(0.027)	(0.020)	(0.025)
Minerals	0.033	0.059	0.096*	0.102´	-0.025**	-0.021	-0.038**	-0.021
	(0.021)	(0.036)	(0.041)	(0.057)	(0.008)	(0.013)	(0.009)	(0.013)
Population	-0.024	-0.121	-0.093	-0.190	0.031´	0.061´	-0.018	-0.042
	(0.057)	(0.115)	(0.089)	(0.126)	(0.016)	(0.030)	(0.018)	(0.031)
Dens2	-0.072	0.123	0.340*	0.457*	-0.078**	-0.108*	-0.052	0.022
	(0.081)	(0.172)	(0.138)	(0.203)	(0.024)	(0.043)	(0.033)	(0.054)
Extdens3	-0.098´	-0.034	-0.280´	-0.181	-0.059	-0.096	0.063*	0.068
	(0.057)	(0.167)	(0.153)	(0.255)	(0.036)	(0.071)	(0.030)	(0.050)
Agri*Prod90	-0.028**	-0.011	0.022	0.029	-0.007*	-0.012	-	-
	(0.007)	(0.019)	(0.016)	(0.023)	(0.003)	(0.007)		
Gas&Oil*Prod90	0.002*	0.001	-	-	-	-	-	-
	(0.001)	(0.002)						
Forest*Prod90	-	-	-	-	-0.006**	-0.007*	-	-
					(0.002)	(0.003)		
Minerals*Prod90	-0.003´	0.001	0.009*	0.013*	-0.002**	-0.003**	-0.001	0.000
	(0.002)	(0.004)	(0.004)	(0.006)	(0.001)	(0.001)	(0.001)	(0.002)
Pop*Prod90	-	-	-0.009*	-0.011´	0.002**	0.003´	0.002*	0.000
			(0.004)	(0.007)	(0.001)	(0.001)	(0.001)	(0.002)
Dens2*Prod90	0.018**	0.010	-	-	-	-	-	-
	(0.004)	(0.010)						
Extdens3*Prod90	0.013*	0.004	-	-	-0.004	0.003	-0.012**	-0.013*
	(0.006)	(0.017)			(0.003)	(0.006)	(0.003)	(0.006)
Prod90	0.025**	0.052**	0.099**	0.120**	-0.003	-0.012**	-0.002	0.008
	(0.007)	(0.010)	(0.013)	(0.014)	(0.003)	(0.004)	(0.004)	(0.005)
N	714	42	714	42	714	42	714	42

(1), (3), (5) and (7) Arrellano (1987) type robust standard errors in parentheses.
(2), (4), (6) and (8) MacKinnon and White (1985) type robust errors in parentheses. Dep. var.: log of mean prod over time
Significance levels: . <0.1 * <0.05 **<0.01.

Table A4. The estimates of the model in the 4th column in Tables 2-5, included milit and univ variables.

	(1) Prod	(2) R&D	(3) Cg	(4) Edu	(5) Prod	(6) R&D	(7) Cg	(8) Edu
Time-varying exogenous variables								
Milit	-0.032	0.116**	0.022	-0.113**	-	-	-	-
	(0.023)	(0.041)	(0.013)	(0.016)				
Time-varying endogenous variables								
R&D-intensity	0.160**	-	-0.093**	0.129**	0.151**	-	-0.096**	0.134**
	(0.021)		(0.013)	(0.015)	(0.021)		(0.013)	(0.016)
Share of c-goods	-1.418**	-0.763**	-	-0.577**	-1.445**	-0.756**	-	-0.635**
	(0.061)	(0.104)		(0.039)	(0.060)	(0.104)		(0.041)
Education	0.059	0.756**	-0.406**	-	0.081	0.693**	-0.420**	-
	(0.051)	(0.086)	(0.028)		(0.049)	(0.083)	(0.027)	
Time-invariant exogenous variables								
Agri	0.271**	-0.190	0.014	-0.056	0.281*	-0.099	0.015	-0.041
	(0.075)	(0.154)	(0.032)	(0.039)	(0.113)	(0.155)	(0.042)	(0.033)
Gas&Oil	-0.003	-0.067**	0.003	0.024**	-0.002	-0.039´	0.005	0.018**
	(0.010)	(0.020)	(0.004)	(0.005)	(0.016)	(0.021)	(0.005)	(0.005)
Forest	-0.045	-0.104	-0.024	0.018	-0.049	-0.044	-0.016	-0.002
	(0.038)	(0.092)	(0.018)	(0.026)	(0.059)	(0.093)	(0.024)	(0.023)
Minerals	0.030	0.082´	-0.023**	-0.034*	0.026	0.044	-0.026*	-0.038**
	(0.021)	(0.046)	(0.008)	(0.014)	(0.032)	(0.047)	(0.011)	(0.012)
Population	-0.003	-0.094	0.023	-0.022	0.008	0.048	0.033	-0.027
	(0.050)	(0.108)	(0.020)	(0.030)	(0.079)	(0.114)	(0.028)	(0.027)
Dens2	-0.095	0.304´	-0.064*	-0.036	-0.117	0.104	-0.075´	-0.054
	(0.072)	(0.157)	(0.029)	(0.046)	(0.113)	(0.166)	(0.039)	(0.041)
Extdens3	-0.162*	-0.261´	-0.058´	0.073´	-0.156	-0.067	-0.038	0.038
	(0.073)	(0.157)	(0.031)	(0.043)	(0.117)	(0.164)	(0.041)	(0.040)
Agri*Prod90	-0.031**	0.021	-0.005	-	-0.032**	0.010	-0.006	-
	(0.008)	(0.015)	(0.003)		(0.012)	(0.015)	(0.004)	
Gas&Oil*Prod90	0.002*	-	-	-	0.002	-	-	-
	(0.001)				(0.001)			
Forest*Prod90	-	-	-0.006*	-	-	-	-0.007*	-
			(0.003)				(0.003)	
Minerals*Prod90	-0.004´	0.009*	-0.001´	-0.001	-0.004	0.002	-0.002´	-0.001
	(0.002)	(0.004)	(0.001)	(0.001)	(0.003)	(0.004)	(0.001)	(0.001)
Pop*Prod90	-	-0.010*	0.001	0.003*	-	-0.006	0.002	0.002´
		(0.005)	(0.001)	(0.001)		(0.005)	(0.001)	(0.001)
Dens2*Prod90	0.019**	-	-	-	0.019**	-	-	-
	(0.003)				(0.005)			
Extdens3*Prod90	0.012´	-	-0.005´	-0.013**	0.013	-	-0.007´	-0.010**
	(0.007)		(0.003)	(0.004)	(0.010)		(0.004)	(0.004)
Prod90	0.030**	0.102**	0.003	-0.010*	0.027*	0.032	-0.002	-0.001
	(0.005)	(0.013)	(0.003)	(0.004)	(0.013)	(0.021)	(0.005)	(0.005)
Univ	-	-	-	-	0.023	0.438**	0.035	-0.061*
					(0.069)	(0.103)	(0.024)	(0.026)
N	714	714	714	714	714	714	714	714

(1)-(4) Ameniya and MaCurdy (1986) standard errors in parentheses.
(5)-(8) Arrellano (1987) type robust standard errors in parentheses.
Significance levels: . <0.1 * <0.05 **<0.01.