Exploring the knowledge space

How national research portfolios evolve over time and space

Master thesis | Innovation Sciences
Copernicus Institute of Sustainable Development
Department of Innovation, Environmental and Energy Sciences
Faculty of Geosciences | Utrecht University
July 20th, 2018

Author: Florian Sloots
4008693
d.f.sloots@uu.nl

Supervisor: Dr. Gaston Heimeriks
g.j.heimeriks@uu.nl

Second reader: Prof. Dr. Bernhard Truffer
Bernhard.Truffer@eawag.ch

Utrecht University

Summary

Knowledge creation is argued to be one of the most important activities of todays knowledge based societies. Fact is that knowledge creation and accumulation is remarkably variable across space and time, and concentrated in relatively few places. Many researchers have studied and tried to explain this variance in the differing ability of countries to develop innovative knowledge and competitive advantage, but still a lot is unknown about the characteristics of the conceptual scientific knowledge space. This research aims to explore trends and correlations in variance of national research portfolios, differences in growth rates of scientific disciplines and connect those to enabling factors and economic impacts. This was done with an explorative quantitative research design, performing analyses with publication data from SCImago JR and additional economic and governance data. Specific correlations were tested using panel linear model regression models with fixed effects to isolate the effect of country and year omitted variables. It was found that distinguishing research portfolios in terms of knowledge complexity, diversity, specialization, unrelated and related diversification and various other characteristics can help to understand knowledge development variations and trends. Knowledge complexity can be linked to stages in economic development. Furthermore the role of internal dynamics of scientific disciplines was confirmed as systematic differences exist in their development and occurrence in the portfolios or different types of countries. Lastly the results provide interesting insights in the role of governance quality and institutions in determining growth potential and as enabling or restricting contextual factors. The findings confirm that general notions from theories on knowledge development also apply in national research trends, such as related diversification and path dependence. Furthermore similarities are found with the literature on economic complexity and varieties of capitalism. Exploring knowledge development through the networked character of publication data proves to provide theoretically interesting results, leads for future research and societally relevant insights.

Acknowledgements

This thesis research has not been written and performed without overcoming a number of obstacles and detours. I want to thank those who have guided me during the process and helped me create a final version of which I can be proud.

First of all I want to thank my supervisor Dr. Gaston Heimeriks for his guidance, useful feedback and interesting discussions. I would like to thank Deyu Li for his advice on data transformations, network analysis and R scripts, which has provided me with useful insights and saved me a lot of time. Furthermore I would like to thank Chris Eveleens for his advice in the approach of the data collection, and Dr. Jarno Hoekman and Dr. Maryse Chappin for taking the time to discuss an appropriate methodology and proper interpretation of results. Lastly, my thanks goes out to my fellow students which have provided me with feedback and support.

Table of content

1. Introduction

Knowledge is an increasingly important resource in todays 'knowledge-based economy', which has also increasingly been acknowledged and included in economic models during the past decades (Cooke & Leydesdorff, 2006). Different types of research have aimed to explain how knowledge develops over time and under what circumstances competitive knowledge is best created. While countries can accumulate wealth by extracting natural resources, it is seldom enough to support long-term sustained economic growth (Soete et al., 2015). Knowledge does provide this potential, as it leads to new discoveries, innovations and technological development (Patelli et al., 2017), which stimulate long-run economic growth and development (Howells, 2002; Balland & Rigby, 2015).

However, in terms of knowledge production the world is spiky; it is very unevenly distributed over regions (Florida, 2005), concentrated in a handful of countries (Petralia et al., 2017), and many countries struggle to replicate levels of productivity and innovativeness of leading regions (Heimeriks & Balland, 2015). As stated by Rigby (2015, p.1924) '…*the empirical understanding of the geography of knowledge amounts to little more than inventories of activity within different economic categories across different locations'*. This underlines the need for systematic understanding of this variation beyond merely summing up differences; looking for patterns and mechanisms.

Therefore, the question is; what determines variation between countries and the concentration of certain types of knowledge in specific locations? The explanation could lie in different aspects of knowledge development. One way to explain variations could the local context of knowledge creation. For instance Soete et al. (2015) illustrate the link between geopolitical events, different national contexts and the productivity of countries in terms of scientific knowledge production. In line with this view, the inequality in knowledge production can be linked to income gaps between nations (Hausmann et al., 2014), as competitive advantage and welfare are determined by knowledge stored in economic, social and organizational networks, and the capability to continuously develop new, complex and valuable knowledge (Petralia et al., 2017; Hidalgo & Hausman, 2009).

Another explanation could lie in the internal dynamics and structure of different types of knowledge which cause some fields to grow faster or develop differently than others. However, besides a few pioneering works, there are no studies on differences in growth rates (Bonaccorsi, 2008). This view is in line with theories on path- and place dependence, which state that knowledge development depends on an existing knowledge base. Combining the two views, knowledge development can be seen as a co-evolution process in which local context and actors influence knowledge development and are affected by it (Cano-Kollmann et al., 2016). Also Bonaccorsi (2008) finds that different institutional contexts enable different types of knowledge production.

Another study trying to provide insight in the connection between knowledge creation and economic impact is the Economic Complexity Index (ECI) developed by Hidalgo & Hausmann (2009). The ECI value, which is based on export data, reflects the complexity of knowledge that is required for a country to produce a specific set of products. The index proves to be a better predictor of economic growth than others such as the six Worldwide Governance Indicators (WGIs), which capture a countries' governance and institutional quality (Hausmann et al., 2014). The value and complexity of knowledge in this sense are determined by the networked character. Different types of knowledge are more related to each other than others and as such certain combinations of knowledge provide more potential for new combinations.

By combining the ECI with the concept of the product space, Hartmann et al. (2017) illustrate that different national product portfolios correspond to different stages in economic development and that pathways of development can be shown over time through the product space. It was found that stages in development can also be linked to different knowledge development strategies; low-income countries usually develop related low-complexity knowledge, but need to develop more complex knowledge in unrelated fields somewhere along the way in order to develop further and catch up with high-income countries (Pinheiro et al, 2018).

However, as Ivanova et al. (2017) indicate, the assumed productive knowledge underneath product portfolios is a latent dimension in the ECI, as it is not directly measured. For this reason they developed *"a Patent Complexity Index (PatCI) on the basis of a matrix of nations versus patent classes",* with the aim to explicitly measure the technological capabilities of countries (Ivanova et al., 2017, p.1). This index does however not succeed to reproduce a similar correlation with economic effects.

While both patents and export data provide valuable insights in the geographical variation of knowledge, questions remain on how countries accumulate and develop knowledge, and how they can use this to climb the ladder of technological development (Petralia et al., 2017). The field of scientometrics can offer a more direct reflection of the process of knowledge creation and accumulation in the form of publication data. This is an advantage as publications data can provide a direct reflection of the knowledge that is produced in a country. This is because the field of scientometrics documents how the research and education system accumulates verified knowledge in the form of peer-reviewed scientific publications over time (Small & Upham (2009).

As stated by Vick & Nagaro (2018) scientific knowledge supports the creation of a knowledge based economy as it generates knowledge that is important for extending existing and creating new economic activity. Furthermore, the study of Patelli et al. (2017) shows that national scientific knowledge development, indicated by scientific publications, also has a positive influence on discoveries, innovations and technological development. Accordingly, Klavans & Boyack (2017, p. 1) state: *'Research portfolio analysis should be a key activity for all stakeholders in the current science system.'* They explain that gaining insight in research portfolios and specific research areas is vital, as the potential strategy and policy choices in science systems are currently not well understood. Studying these type of questions can provide insight in the strengths and weaknesses of the research efforts of a nation, and as such form an important tool for science policy (Nederhof & Van Wijk, 1997).

As discussed by Carayannis & Grigoroudis (2016) it becomes increasingly important for countries to apply smart specialization; to remain competitive it is necessary to adjust scientific knowledge production, emerging opportunities and specializations with current business needs on regional and country level. This is in line with the developments described in the recent Science report by Soete et al. (2015) which describes that countries around the world increasingly invest in knowledge development and employ science and technology strategies.

Examples of studies that provide insight in global patterns in scientific knowledge developments are for instance Horlings & van den Besselaar (2011) who use publication data to show that countries can be clustered on the topics in which their scientific knowledge portfolio is specialized, and that these clusters can be used to explain differences in growth over time. However, Van Elk et al. (2015) find that there is no uniform relationship between scientific knowledge development and economic productivity, which suggests that there is no one road to success and that country specific context matters.

Making use of the networked character of publication data, this study aims to explore the dynamics of national research portfolios over time in a so-called 'knowledge space' as defined by Heimeriks & Balland (2015). That is; this research aims to go beyond the earlier mentioned 'inventories of activities' to explain variety of scientific knowledge development over space and time. This will be done taking into account potential enabling or restricting factors, both internal and external to knowledge development. Furthermore the relation with economic growth and different stages of development will be explored. In order to do so the following research question will be employed:

How do national scientific research portfolios vary over time and space and how can this be explained?

In order to answer this question, this study will take an explorative approach. Using international publication data from 1996 until 2016, the connection between geographically varying knowledge development, and development of specific research topics will be studied. This will be combined with other databases in order to explore the co-evolution or two-way interaction with supporting factors in the local context and possible economic impact or outcomes.

As stressed by different authors, insight in knowledge development in science systems is important to inform strategic choices. Before knowledge development and its outcomes can be influenced the interaction between local context and internal dynamics should be better understood. More insight in these patterns can better inform which contextual factors and search regimes are suitable to stimulate economic development through scientific knowledge creation.

This can help countries in different positions on the ladder of economic development to gain insight in their current knowledge base, their respective position, and what opportunities lie in the vicinity of related knowledge of their portfolio. This can be used to make informed decisions on the possible specializations and related diversifications which may be most fruitful to develop. This is in line with the developments that smart specialization becomes increasingly important in todays knowledge society to create and maintain competitive advantage.

Besides practical managerial, policy and governmental applications, this study also provides novel scientific results. This is done by combining ideas and theories from different strands of relevant scientific literature on knowledge development such as evolutionary economic geography, scientometrics, econometrics and innovation sciences. As the cause of different growth rates in scientific fields in different locations has not been studied extensively yet, exploring and describing patterns can be a valuable addition to literature.

Lastly, by taking a network approach to analyze publication data, as opposed to patent or trade export data, this study can provide new results and make use of new methods in the field of scientometrics. Also by employing new methodologies and network measures that can provide more insight in the dynamics of the content of national scientific portfolios, the varieties that exist over time and space, this study may provide an example and inspiration as well as leads for future research.

This report is structured as follows. Section 2 provides an overview of relevant literature and creates a theoretical framework that will be used in this study. In section 3 an elaborate and precise description and motivation of the research design and methods are presented. Section 4 presents the results of the exploration and data analysis in order to answer the research question and proposed hypothesis. Finally section 5 provides a discussion of the most important implications, limitations and suggestions for future research, ending with a conclusion summarizing the answers to the research question.

2. Theory

This chapter provides an overview of relevant theory, existing ideas and gaps in literature and fits them together into a theoretical framework that will help explore patterns and trends in scientific knowledge development in this study. Important concepts are defined and expectations are formulated in four main hypotheses that will be tested. First an overview will be provided of relevant theories on the two dimensions of scientific knowledge development in countries and on development of specific scientific disciplines.

National research portfolio

As stated in the research question this study focuses on scientific knowledge development in national research portfolios. The first important concept *knowledge development*, is defined as the creation and accumulation of scientific knowledge in terms of peer reviewed publications. The other concept is the *national research portfolio*; which represents the composition different disciplines in the set of scientific knowledge output a country produces.

The complexity of scientific research is illustrated by Gómez-Núñez et al. (2011, p.742): *"Nowadays, research is influenced by factors such as its strong relationship with society, the ultra-specialization of areas and disciplines of knowledge, the competitiveness exercised by increasing practitioners, groups and research institutions, or the dynamism".* While knowledge production can be aggregated and evaluated on different levels; from personal, organizational, to cities, regions, countries and global scale, as it is a complex process involving heterogeneous actors from different levels and areas of the science system, the nation level provides a means of comparing outcomes of this complex process (Horlings & van den Besselaar, 2011). This can also be motivated by the fact that country specific characteristics matter for knowledge development. As discussed by Boschma & Capone (2016) specific institutions can have a stimulating or restricting effect, but also capabilities and learning capacity differ per country and affect their productive capacity.

Furthermore, the nation level facilitates comparison of economic effects and other characteristics, as country level data is widely available and many studies aim to explore or explain differences and similarities between countries. It would for instance be more complex to take into account local context and allocate the precise economic impact on a university or research institute level. Therefore this study also focuses on research portfolios on the nation level.

Scientometrics and scientific disciplines

The composition or content of national research portfolios plays an important role in this study. This refers to the different types of research, research areas and fields, or scientific disciplines that make up the subset of all scientific knowledge that a specific country produces. The study of the dynamics of disciplines in the production of scientific literature falls under the field of bibliometrics, informetrics and scientometrics. The latter was used by Nalimov in the 60's, referring to the study of science, growth, structure, interrelationships and productivity (Hood & Wilson, 2001).

There are quite some studies that use national publication data to study how countries develop or how countries compete with each other such as: Kharabaf & Abdollahi (2012) on science in Iran, Gholizadeh et al. (2014) comparing ASEAN countries with the top 10 countries in the world, Mêgnigbêto (2012) on countries in West-Africa and Bashir (2013) on science in Pakistan. What these studies have in common is that they mainly use absolute publication and citation counts of countries to reflect the productivity of their science systems.

While these insights can be useful, there are studies that dive further into the content of research , the way that disciplines develop and interact, and make better use of the rich possibilities of publication data. This study aims to use measures that provide more insight than just aggregate amounts and aims to provide more insight in differences in portfolio content over time and space as well.

For instance Guerrero-Bote et al. (2007) argue that there are differences between scientific disciplines in the extent to which they import or export knowledge to and from other disciplines, indicating that some may be more multidisciplinary or related to other disciplines that others. This suggests that some disciplines would be more likely to develop in the same location than others.

The development of different scientific disciplines has also been studied by Heimeriks & Balland (2015) who find that there are fundamentally different characteristics of different disciplines. They conclude that this has important implications for accumulation and relatedness of knowledge and should be taken into account in smart specialization strategies. Bonaccorsi (2008) also pays attention to specific disciplines in science, making a distinction between traditional sciences and new sciences such as information, materials or life science, which grow a lot faster and develop in more diverse directions.

The knowledge space

The concept of the *knowledge space* plays an important role in this study. It has been described by Heimeriks & Balland (2015) as a space in which each location only comprises a small subset of all possible recombinations of scientific topics, and searching through the space brings costs with it.

The concept builds on the theoretical concepts of path- and place dependence and related diversification, and the fact that knowledge development is a cumulative process. Important notions from evolutionary economics that knowledge development is place dependent and path dependent (Heimeriks & Balland, 2015; David, 1994). Knowledge is often complex and tacit in nature; stored in personal, interpersonal and organizational experience and networks, and thereby bound to geography (Petralia et al., 2017). This makes spatial proximity important to lower barriers and costs of transmission (Breschi et al., 2003) and to gain the possibility to access or to profit from knowledge spill-overs (Sorenson et al., 2006).

Furthermore, specific knowledge is developed building on an existing knowledge base, as path dependence implies that new knowledge evolves by recombining existing knowledge building blocks (Arthur, 2007). The existing knowledge base, and underlying skills and capabilities make that it is easier to specialize or diversify in knowledge that is similar or proximate, and costly to search and explore new types of knowledge that are more distant (Breschi et al., 2003).

Related to path- and place dependence is the distinction between related and unrelated variety in knowledge as made by Frenken et al. (2007). What follows is that a portfolio that contains a variety of related knowledge provides the possibility for knowledge spill-overs and recombinations into new types of knowledge. This way the current knowledge base determines the possibilities to certain pathways of development into related fields, and complicates or restricts the access to unrelated fields. This is related to the finding that countries are more likely to develop new activities in fields that are similar to their current activities, also referred to as 'branching' (Hidalgo et al., 2007). This relatedness concept is also observed for development of research areas (Guevara et al, 2016), as illustrated for instance by Boschma et al. (2014) who find that the emergence of new scientific topics in biotech happens in cities where related knowledge already exists.

This is related to the concept of the product space, based on the relatedness of export products, calculated from the probability that they co-occur in different countries (Hidalgo et al., 2007). They explain this concept as a *"network of relatedness between products"* where "… *more-sophisticated products are located in a densely connected core whereas less- sophisticated products occupy a less-connected periphery"* (Hidalgo et al., 2007, p.482). This implies that certain products are more sophisticated and complex as they provide more adjacent possibilities, the chance to diversify into related fields, than other fields would.

Similar to this concept a knowledge space can also be constructed. In this space a network of research disciplines is positioned based on their relatedness. In this network it can be defined what the respective position and development of countries is based on their research portfolio. The position and movement of countries and scientific in this space forms the base of the theoretical framework of this study. Expectations on knowledge development are related to characteristics of this space and related theories from relevant literature.

Horlings & van den Besselaar (2011) find that there is a linear trend between absolute scientific output of countries and the varieties of areas in which they publish. This means that smaller countries (in terms of publication count) are more specialized in a few areas, while bigger countries have a more diversified portfolio. Following the characteristics of the product space, and the notion that technological diversity is thought to be beneficial as this leads to a higher number of possible recombinations (van Rijnsoever et al., 2015), this may also be true for scientific knowledge development.

So it can be expected that there are countries with a specialized, smaller portfolio with less valuable or complex knowledge, with less possibilities for related diversification, in the periphery of the knowledge space. Other countries may have bigger more diversified portfolios with more complex knowledge and more related diversification possibilities. This would imply that there is a development trajectory from the periphery to the centre that countries can go through by entering new fields and broadening their portfolio. Increasing publication output in the same fields; specialization, or related diversification in fields closely would not move them much further to the centre of the knowledge space, on the other hand unrelated diversification might help increase variety in their portfolios and increase the average complexity and related diversification possibilities of their knowledge.

Economic development

This distinction of countries between the types of knowledge in their portfolio in the extent that they are sophisticated or complex has implications. The economic complexity index of Hausmann et al. (2014), based on the complexity of its export products and the underlying productive knowledge, has proven to be a good predictor of a nations economic growth. Cicerone et al. (2017) find that a better position in the product space network leads to higher economic outcomes.

Not only position, but also movement in the space seems to have implications. Pinheiro et al. (2018) find that the type knowledge development is linked to different stages of economic development; low-income countries develop into related low complexity fields, and high-income countries develop into related high complexity fields. Similarly, Petralia et al. (2017) link different stages of development with ‘climbing the ladder of technological development’ towards more complex and valuable technologies.

Based on these notions of the knowledge space, related and unrelated diversification and ideas from earlier discussed research expectations on the link between countries knowledge development, position and movement in the knowledge space and economic development can be formulated. These expectations are defined in the following hypotheses;

H1: countries with a portfolio containing complex knowledge are positioned in the centre of the knowledge space network, countries with low complexity knowledge are positioned in the periphery.

H2a: Most knowledge development by low-income countries is in related fields of low complexity knowledge in the periphery of the knowledge space. Most knowledge development by high-income countries is in related fields of high complexity knowledge in the centre of the knowledge space.

H2b: Developing countries move from developing low complexity knowledge, to developing high complexity knowledge through unrelated diversification, and thereby move from the periphery towards the centre of the knowledge space.

How the concepts from the first two hypotheses are expected to relate to the knowledge space is depicted in figure 1.

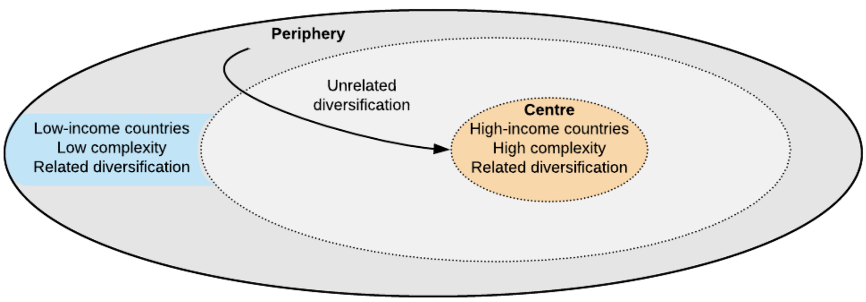


Figure 1. - Knowledge space

While complexity may reflect a variety of different types of knowledge in a portfolio, providing some insight in the content, specific disciplines might also play a role. This is in line with the results of Hartmann et al. (2017) who link different product portfolios to different stages in economic development.

It can be expected that the same is true for research disciplines. Horlings & Van den Besselaar (2011) find that countries can be clustered based on specific fields of research in their portfolio. This is also in line with the findings of Heimeriks & Balland (2015) who find that different disciplines offer different opportunities for related development. Thus moving to one area may be more profitable than moving to another, and different areas may play a different role in different stages of development, depending on the position of a country in the knowledge space and the content of its knowledge base.

Combined this may suggest that specific research topics play a role in development paths of nations. Therefore it can be expected that besides complexity of a knowledge portfolio, also specific topics matter. Countries may thus move into areas that increase their chance on economic development. Research areas that offer more opportunities for economic impact through innovation and the creation of new sectors in a countries economy may help them to 'climb the ladder' and provide them with more means to invest in new knowledge development. Based on this expectation the following hypothesis is proposed:

H3: Different stages in development, and nations with different levels of income are linked to different compositions of their research portfolio.

Bonaccorsi (2008) also pays more attention to those specific disciplines in science, making a distinction between traditional sciences and new sciences such as information, materials or life science, which grow a lot faster and develop in more diverse directions. Connecting this to country context, studies have argued that different types of institutions have the ability to enable or constrain different forms of economic activity (Jackson & Deeg, 2008). This could also be expected to be true for scientific development. For instance some institutional settings, such as in the US, seem to be able to enable growth of emerging sciences better than others, such as those in Europe (Bonaccorsi, 2008).

Fu et al. (2011) also stress that the benefits of international technology diffusion can only be grasped by countries that have the right indigenous modern institutional and governance structures to facilitate an efficient innovation system. This is related to the earlier mentioned learning capabilities of a country. Institutional context may play an important role in facilitating knowledge development. Furthermore, Soete et al. (2015) illustrate the effect geopolitical context and events can have on scientific knowledge production.

Other studies illustrate the influence of institutional context on product portfolio (Hartmann et al., 2017) and on national comparative advantage, by comparing varieties of capitalism (Hall, 2001). The idea is that countries can be categorized in different types of economies based on their institutional and governance structure. The ability of countries to facilitate radical innovation versus incremental innovation can be linked to the extent to which institutions coordinate the economy.

On the one hand Witt & Jackson (2016) illustrate that low coordination in all institutional spheres except employment relations; defined as liberal market economies, seems to facilitate radical innovation better, and that coordinated institutions, defining coordinated market economies, can be linked to incremental innovation. In terms of knowledge development this can be linked to either developing in incremental steps in the same or similar fields; specialization or related diversification, or on the other hand diversifying into unrelated, radically new fields or areas of knowledge; radical innovation. However, Witt & Jackson (2016) also stress that there is limited empirical evidence for this hypothesis comparing liberal market economies and coordinated market economies, as proposed by Hall (2001).

These ideas could be used to further specify the relation between knowledge development and countries in terms of enabling or restraining factors for successful knowledge development in specific scientific areas. Also it could be expected to be a two way influence. While a good environment may enable knowledge development, the positive impacts of knowledge development on the economy and society may have a positive feedback on the governance and institutions of that country. Following these ideas this study will explore what relation exists between institutional context and knowledge development, according to the following expectation:

H4: The institutional, governance and economic context of a country has influence on the composition of its portfolio in terms of specific types of knowledge and their respective growth.

In the next chapter, the proposed theoretical framework are further operationalized. Furthermore methods are discussed to test the four proposed hypotheses.

3. Methods

In order to answer the research question and to test the proposed hypotheses, this study employs a quantitative research design. The specifics of the research design, including methods of data collection, operationalization and data preparation and analysis are discussed in this chapter. Lastly an explanation is provided on how this study aims to provide a proper research quality.

3.1 Research design

This study employs a quantitative research design which contains both explorative and deductive elements. An explorative approach was chosen deliberately in this research. This was done in order to account for the fact that the both research question and intended data and methods to answer this question are part of an emerging field of science. The aim is not solely to test expectations from different related strands of literature, but also to keep an open attitude to unexpected trends in the data.

In order to provide insight in the dynamics of knowledge development over time, a longitudinal design was chosen. This provides advantages over a cross-sectional design; such as that it enables the researcher to infer that some effects occur after changes in independent variables (Bryman, 2012). This does not ensure causality, but can provide more insight in potential causality than a cross-section does. In this case a longitudinal design enables the identification of sustained trends in the data over time.

The first part consists of an explorative analysis of the database, which was performed in order to explore patterns and find possible correlations between variables, such as general characteristics, interrelations and dynamics of scientific knowledge portfolios and scientific disciplines. This was an iterative process, meaning initial results were re-evaluated and used for further analysis steps. In order to further explore found correlations in the analysis and study possible implications, theoretical literature was used to help explain and interpret findings or to generate ideas for additional quantitative analysis on specific subsets of the data, or to categorize entities in the data.

This was combined with the deductive part of the study, in which correlation regressions were used to analyze the data. This was done with the aim to find out whether the expected trends and relations between variables in de database exist, such as proposed in the hypotheses, and to explore whether there are unexpected new correlations or trends that prove to have significant effect.

The unit of analysis in this study is mainly the nation. The unit of observation is the national research portfolio reflected by publication data. As will be further elaborated on in the data collection section; a sample containing almost all entities in the population of countries will be used. This is strived for, however, data on some external variables the data covers a lower amount of countries, limiting some aspects of the analysis, including variables from these data sources, to a lower amount of countries. Besides nations, the focus also lies on fields or areas of scientific research in some parts of the data analysis.

Lastly for some analysis subsets of the data are further explored, as these can be interesting outliers, or cases that form an example for general trends or patterns found in the analysis. This is done to illustrate how insight in knowledge dynamics can be used to explore trends in knowledge development and possibly inform policy and strategy decisions.

3.2 Data collection

Scientometric data

In order to explore the dynamics of knowledge developments in national research portfolios, data on the characteristics and scientific output of science systems over a wide set of countries and over a longer time span is required. Furthermore, to provide insight in internal dynamics of scientific disciplines, also data on specific fields in science is necessary.

Large databases that provide access to large amounts of scientometric data are Google Scholar, Scopus, Web of Science (WoS), and there are databases more specialized in specific fields such as PubMed which contains citations of biomedical literature (Falagas et al. 2008a). Scopus provides advantages over other databases in that it has a higher coverage of journals and a 20% higher coverage of citation data than Web of Science. Google Scholar provides access to a larger scope of literature, but it is stated to provide results with inconsistent accuracy and citation data can be inadequate (Falagas et al., 2008a).

As stated by López-Illescas et al. (2009) Scopus provides more coverage on nationally oriented journals than WoS. Furthermore, after 1996 when citation analysis is provided, Scopus outperforms WoS due to the availability of the breadth of the database (Moed, 2002; Powell & Peterson, 2017). While the amount of journals originating from non-English languages in Scopus is still underrepresented in the database, constituting 15% of the total database, SJR is still argued to be a better option to provide and estimation of the value of these journals than other databases (Falagas et al., 2008b).

Biases exist, both in the underrepresentation of English-language journals compared to other languages, and in the the representation or share that Natural Sciences, Engineering and Biomedical research take in compared to Social Sciences and Arts and Humanities (Mongeon & Paul-Hus, 2016). This means that caution is required with interpretation of absolute amounts of publications and comparisons between countries and disciplines. The consequences of the choice for this data for the interpretation of the results will be elaborated on in the discussion chapter.

Comparisons and reviews of databases show that no perfect database exist. It can be argued however that the coverage and quality of Scopus are favorable. While the Scopus database is not freely available, SCImago Journal Rank (SJR) provides public access to part of the publication data from Scopus on the SCImago Journal & Country Rank portal. SCImago is *'a research group from the Consejo Superior de Investigaciones Científicas (CSIC), University of Granada, Extremadura, Carlos III (Madrid) and Alcalá de Henares, dedicated to information analysis, representation and retrieval by means of visualization techniques'* (SCImago, n.d.). The database provides journal and country scientific indicators based on information provided on Scopus® by Elsevier B.V. As the SJR has proved in multiple studies to be useful for analysis or comparison of knowledge development in specific countries or world regions or scientific disciplines (Kharabaf & Abdollahi, 2012; Gholizadeh, 2014 & Zacca-González et al., 2014) and the underlying Scopus data can be argued to provide sufficient coverage and quality, this database will be used as the main data for this study.

The SJR data to be used is publication data of 239 countries worldwide, categorized in 27 major thematic areas and 310 specific subject categories, the same as used in Scopus, for the period of 1996 up 2016. The categorization is based on co-citation clustering using citation data from over 21,500 journals from more than 5000 international publishers, (SCImago, n.d.). This means that for every country and year a number of publications is provided with additional information on citations and the type of publications that have been produced.

An explanation on the metrics behind the publication database and the used categorization and sub-categorization can be found in the article by Gómez-Núñez et al. (2011), more information on the journal indicator of SJR can be found in the article by Guerrero-Bote & Moya-Anegón (2012). It should be noted that periodic updates of the SJR portal which include retrospective data will make that complete replicability is not possible through repeated data collection. However, on request the used data for this study can be made available.

In order to enable fast, systematic and precise data collection, the database was retrieved, restructured and stored using R (R Core Team, 2013) running in RStudio (R Studio Team, 2015). This reduced the chance of human error in the collection of the high quantity of data. Further processing, analyses and visualization of the data was done making use of amongst others R Studio and Microsoft Excel.

Complementary data sources

As this study takes the nation level as unit of analysis, there is a rich pool of data sources publicly available, on different types of country characteristics. As described in the theory chapter, both enabling or restricting circumstances as well as economic effects on country level are expected to be related to knowledge development. In order to explore this expected connection, data on country characteristics has to be collected to measure concepts from the proposed hypotheses.

For some economic and institutional characteristics, quantitative data was collected from publicly available sources such as The World Bank database. The World Bank provides access to numerous data bases such as the World Development Indicators (WDI) database and the World Governance Indicators (WGI). This was specifically used to collect data on indicators such as GDP or investment in R&D from the WDI database. The economic data originates from the World Bank national accounts data, and OECD National Accounts data files (The World Bank, 2018a). Information and data on institutional setting was collected via The World Bank (2018b) from the WGI database, as developed by Kaufmann et al. (2010). Furthermore qualitative data on institutional characteristics was retrieved from scientific literature such as articles on Varieties of Capitalism, different types of economies and business systems from (Witt et al., 2017), which was used as a source to inform categorization of countries in specific clusters.

Lastly, in order to provide a means of comparing the produced indicators in this study to indicators from studies based on other types of data, data on the Economic Complexity Index was collected, based on the research behind the Atlas of Economic Complexity by Hausmann et al. (2014).

3.3 Operationalization

In order to explore and test for trends and correlations in the data, the concepts from the theory chapter had to be operationalized into indicators using the collected data. This is done both for country or portfolio level concepts as well as some concepts relating to processes or scientific disciplines.

The first concept that is central to this study is the *national research portfolio,* of which several characteristics will be operationalized using different measures based on the amount and type (in different categories and subcategories) of publications a country produces in a year. The hypotheses include a number of concepts related to national research portfolios that are further defined in table 1.

The first two concepts relate to knowledge development, as they measure change of portfolios over time by comparing two or more years. The next five concepts describe portfolio characteristics at one point in time, but can also be used to track changes in the portfolio over time by comparing years. The concepts are further elaborated on below.

Concept	**Description**	**Measure**
Growth *(process)*	Increase in the amount of publications of a country (in specific fields)	Increase in the amount of publications of a country (in specific fields) over time
Specialize in existing fields *or* diversify into related *or* unrelated new fields *(process)*	A country produces publications to existing, new related or new unrelated research fields.	Percentage of publications in same, related or unrelated fields compared to year before. Relatedness measure EconGeo Package (Balland, 2017).
Portfolio size	The scientific output in terms of publication count per country per year	Number of publications in a portfolio
Knowledge complexity	The complexity of a countries knowledge portfolio	Shannon entropy of the portfolio - EconGEO package (Balland, 2017).
Knowledge ubiquity	How ubiquitous is a specific area or field in the knowledge space, or how ubiquitous is the knowledge in a portfolio on average	Ubiquity measure EconGeo package (Balland, 2017).

Specialized vs. diversified portfolio	The extent to which a portfolio reflects specialization (concentrated in a low number of research fields) or diversification (spread over a high variety of fields)	1) Hoover specialization coefficient - EconGeo Package (Balland, 2017). 2) Nr. of areas or fields present in portfolio compared to total (27 areas, 310 fields)
Impact of knowledge development	The impact of publications in a specific country or field in terms of citations	1) Average amount of citations per document (per country/per field 2) Hirsch index score

Table 1. - Operationalization of concepts related to national research portfolios and knowledge development.

Growth

The growth of a national research portfolio is reflected by the amount of publications, the size of the scientific output, that are produced per country per year. The static concept of *portfolio size* of a country in one year is used to construct the dynamic growth variable. In this regard a higher level of growth equals a higher number of publications added per year, reflecting increasing efforts or accomplishments of a country in knowledge development. When comparing the growth rates of countries over time not the absolute amount added, but the percentile increase is measured. As this is the increase in amount of publications added per year, it is actually the increase-in-growth factor rather than the growth factor. This is due to the fact that the total amount of publications up to 1996 is not known for every country, and thus the cumulative amount can not be used.

Furthermore, when comparing the growth of different scientific areas or fields, growth can be corrected for the average growth of scientific publications worldwide per year in order to look whether the relative increase in growth of a country or field is above or below average, reflecting whether a fields is relatively stable, emerging or decreasing as compared to the global scientific knowledge development

In order to take into account the year to year variation in calculating the above or below average increase in growth rate the following steps are taken. For each research area and field the amount of publications per year are corrected for the average growth of publication output per year. Next, the resulting above or below average increase in growth rate is determined by fitting the amount of publications per year over the whole period to a linear model. The average slope of the line that best fits the trend is chosen as a reflection of the above or below average increase in growth. This was done in order to take into account year to year variations in scientific output, rather than dividing the difference in output between the first and last year by the amount of years.

Knowledge space, specialization, diversification and relatedness

The relatedness between fields of scientific knowledge can be formalized as a network, the knowledge space, similar to the way technological knowledge reflected by patents is conceptualized in Balland et al. (2017). The knowledge space is then represented by an n*n network where the nodes represent different fields of scientific knowledge and the edges, the distance between nodes is determined by their relatedness. Within this knowledge space network, the position and direction of development of a countries portfolio can be determined. Knowledge development is either specialization, related or unrelated diversification.

In order to capture the specialization or diversification in this knowledge space over time, publications in a portfolio of one year are compared to those of the year before. When publications are published in the same field as the year before they fall under specialization. When publications are in different fields than before this reflects diversification, as a country that has developed those has had to invest in the required capabilities to produce this new related or unrelated knowledge.

In order to measure the percentage of related and unrelated diversification a relatedness measure similar to the one used by Boschma et al. (2015) or in Balland et al. (2017). Whether fields are related is determined by a relatedness measure based on the co-occurence matrix of publications in fields vs. countries. The co-occurence matrix thus captures the amount of times a publication in specific fields co-occurs in the same country. When two fields co-occur more often in the same country it is assumed that they are more similar, as they are produced under similar conditions.

The relatedness function from the the EconGeo R Package (Balland, 2017) was used to produce a normalized measure of relatedness for each pair of fields. To test the robustness of the analysis this was done using two different normalization methods. The association method which is more appropriate for this type of research than the cosine or Jaccard measure, as argued by Van Eck & Waltman (2009) was used. And second, an alternative association method, the probability index, as developed by Steijn (2017) was used. The association method reflects whether the found number of co-occurences exceeds the expected amount of co-occurences assuming independence of the different fields of scientific knowledge. When this is exceeded this means that two fields co-occur more than the calculated probability and that they are more similar.

	field	relatedness
16	1204	112.4
17	1205	117.7
20	1208	112.1
25	1213	113.8
49	1409	117.2
203	2744	175.7
222	2906	114.6
226	2910	112.9
234	2919	112.1
237	2922	169.3
262	3302	115.8
280	3321	112.3
296	3603	113.3
303	3610	112.9
306	3613	113.4

Table 2. - Top % related fields

Thus, for each field the relatedness to all other fields can be determined using the co-occurence data from each years publication data. For each field the top 10% most related fields were labeled related, and the rest was labeled unrelated. As a robustness check 5% and 20% were also used as thresholds. When comparing a countries portfolio of year *i* and year *j* it can be determined what percentage of fields or of the amount of publications (both were performed to check robustness of the analysis) in year *j* is in a related or an unrelated field. This way the knowledge development patterns can be measured.

The measure can be illustrated using the portfolios of Afghanistan in 1996 and 1997. In 1996 the portfolio consists of one publication in field 2701, 3303 and 3305. When in 1997 publications are produced in the same fields this is specialization. As described above, we can also determine the 5% most related fields to the three fields in the portfolio of 1996, this is shown in table 2. When publications are added in one of those fields it is labeled related diversification. Other publications are labeled unrelated diversification. The portfolio of Afghanistan of 1997 consists of one publication in field 2701, meaning that knowledge development is 100% specialization in these years. This can be done for each country over a longer period.

Furthermore the measures can also be used to track for each field or area how much of the growth over time is in countries that specialize in the field, enter through related diversification or through unrelated diversification.

In order to test whether the resulting classification of fields as related or unrelated is similar or different than the categorization of scientific fields within scientific areas from the SJR database, an alternative measure was also used. Specialization is in that case determined the same way. When the first two digits of the code for field where the same, meaning that the publication was in the same area as publications the year before, it was labeled related diversification. When publications are in a different field and area than the year before they were labeled unrelated diversification. This method is based on Frenken et al. (2007) who apply this categorization in a study using patents which are also structured in categories and subcategories. This is illustrated in figure 2. Resulting outcomes are further discussed in the results and discussion chapter.

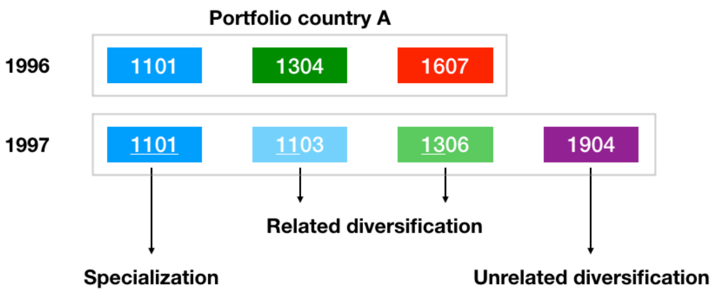


Figure 2. - Labeling knowledge development using the existing categorization.

Knowledge complexity

In order to look further than earlier research, which studied correlations between publication and citation counts and GDP, this study uses a measure that reflects more of the content of a countries knowledge portfolio.

In order to compare research portfolios over time and between countries it is useful to capture the complexity of a set of knowledge in one measure. For this purpose the Shannon-Entropy measure was chosen. While the application of entropy to capture complexity in information content dates back to Shannon & Weaver (1949) the application in economic geography is relatively new. However, as explained and applied in Frenken et al. (2007) using entropy as a measure in the context of diversity or diversification has several advantages, and can be used to provide insight in the variety between different regions.

Entropy captures both the variety and balance of the amount of publications in a portfolio in different fields. Originally entropy captures the randomness complexity, by probabilistically computing the degree of disorganization of a network (Hancock, 2016). It can also be used to determine the inhomogeneity in the distribution of a variable, which corresponds to the earlier mentioned notion that knowledge is unevenly distributed among regions and is concentrated to a large extent to a handful of countries. The original formula of entropy is as follows:

$$\text{Entropy} = -\sum_{i=1}^{n} p_i \log_2 p_i$$

In this specific context the entropy is: - “the sum of the vector of relative frequencies of publications in different fields” multiplied by “the base-2 log of the vector of frequencies” + $1*10^{-9}$, which is an application of the Entropy function from the EconGEO R package (Balland, 2017). The entropy is higher when there is a higher variety of publications in different research fields, and when the total amount of publications is higher. As the entropy of every countries’ portfolio is determined, also the entropy measure both captures variety and amount of publications within portfolios and the differences between different portfolios. This way the relative complexity of a portfolio as compared to another, or the increased complexity as publications are added in a country over time, can be measured.

Portfolio characteristics

Besides the earlier mentioned portfolio size and knowledge complexity, a number of other portfolio characteristics are used as variables. The amount of citations and H index of publications in a portfolio are also used to reflect the scientific impact of different countries over time as provided by the SJR database.

To measure the extent to which the publications in a countries portfolio are concentrated in a small number of scientific areas or fields - specialized - or whether they contain a large variety of publications - diversified - the Hoover specialization coefficient is used. This measure reflects the degree to which a portfolio is concentrated in a few (high specialization) or spread over many (low specialization) scientific fields. An alternative and more simple measure that is also used in the analysis is the nr. of fields in which a country has publications in its portfolio from 1 to 310.

Lastly the ubiquity measure from the EconGeo R package (Balland, 2017) is used to reflect whether fields of science are ubiquitous, and whether portfolios on average contain fields that are more or less common.

Complementary country variables

To test the proposed hypotheses, besides the indicators on research portfolios, also other country level concepts are operationalized. The operationalization of these concepts is shown in table 3.

In order to test the correlation between knowledge development and economic development, economic growth or different stages in development, the log transformation of GDP is used. Besides economic indicators also institutional and political context is taken into account. In order to do this data on governance quality is used reflected by six indicators.

Furthermore, the categorization of types of business systems (Witt et al., 2017) will be used to cluster countries in different types of economies, based on a combination of the literature on varieties of capitalism, and the business systems framework. As the varieties of capitalism approach includes mainly developed countries, this study will take into account a broader scope of institutional structures.

Concept	Description	Measure
Economic development	Welfare in a country in terms of Gross Domestic Product (GDP)	The log transformed GDP per year, or the increase/decrease over time
Institutional context	Different institutional structures that define countries institutions and economies	Types of economies based on Witt. Et al. (2017): *"Highly Coordinated, Coordinated Market, Liberal Market, European Peripheral, Advanced Emerging, Advanced City, Arab Oil-Based, Emerging, and Socialist Economies."*
Governance & politics	Difference in governance and political structures in countries	World Governance Indicators (Kaufmann et al. 2010).

Table 3. - Operationalization of other concepts and control variables.

3.4 Data analysis

The first explorative part of the data analysis consisted of descriptive statistics and exploring trends and correlations in the data through plots, histograms, calculations and map visualizations. For this purpose standard measures and steps were performed as will be presented in the results chapter.

For the deductive part, in order to analyze whether the expected correlations between variables, in the panel dataset that was collected and constructed, multiple statistical regression tests where performed using 'R Studio'. In order to determine the right model for the type of data that was used, different models were tested and evaluated.

For the regression analysis a pooled OLS model, a fixed effects model and a random effects model were estimated. A pooled OLS model can be used to test whether there is a correlation between the variance in the independent variable and the variance in the dependent variable in general. This model treats pools all observations together into one sample and only takes into account differing years of observations, no other objects (Woolridge, 2002). However, as discusses by Cameron et al. (2011) when standard errors should be adjusted for clustering if they are correlated within groups of observations, such as countries. Woolridge (2002) also states that different types of analysis can be more appropriate if different cross-sections have the same structure, such as in panel data.

An advantage of analysis with panel data is that there is a way around the omitted variable bias. When unobserved variables do not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics. By introducing a dummy variable for countries the effect of these country specific variables can be isolated (Oaxaca & Geisler, 2003). As it can be suspected that there are country specific characteristics that influence the relation between knowledge development and economic development, the fixed effects model is deemed to be appropriate in this study. If aggregate trends over time are thought to have an important influence on the outcomes of the model, time fixed effects can also be added to the model to control for the effect they might have.

The random effects model is appropriate in case the hierarchical structure or role of subjects in the data are uncertain or unknown.

Assumption tests

In order to prove which model is most appropriate, assumption tests were performed. An F test was used to determine that the fixed effects model was more appropriate than the OLS model, and the Hausman test was used to decide whether the fixed effects model should be used instead of a random effects model. In case of the fixed effects model an F test was used to test whether time effects dummy variable should be included to control for aggregate trends. This motivated the choice for a fixed effects model with country and year fixed effects. Results of the tests are presented and further elaborated on in appendix C.

Econometric model specification
In order to estimate how different variables related to knowledge development influence economic development, or in order to estimate how other country level characteristics influence knowledge development panel linear models are used. An example of one of the basic econometric equations which is estimated can be written as follows:

$$logGDP_{c,t} = \beta \, entropy_{c,t} + \varphi_c + \alpha_t + \varepsilon_{c,t}$$

In this case logGDP is the dependent variable and entropy the independent variable. However the same model is used for different dependent and independent variables in the analysis. The formula can be explained as follows: The model estimates how economic wealth of a country (c) in a certain year (t) is influenced by on the entropy in that country (c) and year (t). Furthermore, φ_c is a country-fixed effect, α_t is a time-fixed effect, and $\varepsilon_{c,t}$ a regression residual.

The analysis was used in different ways; to study the correlation between portfolio characteristics, variables on knowledge development on economic outcomes, and to study the correlation between enabling or restricting country specific variables on knowledge development.

In the interpretation of the model it is important to note that between country differences are captured by the isolated fixed effects. What the model can estimate is the effect that variance of the independent variables have on the variance in the dependent variable within countries over time. As discussed by Mummolo & Peterson (2018), caution is required in the interpretation of the results of fixed effects regression analysis as many studies report counterfactuals. In order to ensure proper interpretation readers should take note that for all presented regression output tables estimates can be interpreted as follows: “as X changes within countries over time, Y changes. . . ”. In order to strive for proper research quality the checklist provided by Mummolo & Peterson (2018) has been taken into consideration in the interpretation of the performed analyses.

Between estimator
To use the panel data for regression analysis to estimate the difference between countries instead of the difference within countries over time a between estimator model was used. In this model the time variable is left out as for all variables the average over the whole period is used. This model was used to test for the influence of the categorical variable of different types of economies and whether being in one category resulted in significant differences in certain country level knowledge development characteristics as compared to other types of economies.

ANOVA analysis
Furthermore in the analysis clusters of countries were introduced. In order to legitimize the clustering of countries according to, instead of the data itself, an ANOVA test was performed. This was done to illustrate that the variance between countries within clusters is significantly smaller than the variance between clusters.

3.5 Research quality

In order to pursue a high level of research quality, four criteria will be used to assess the quality of this study: *reliability, internal validity, external validity* and *construct validity*. Reliability is reflected by the replicability of the research, yielding the same results, which can be achieved by transparency of the analysis (Yin, 2013). To achieve higher reliability, all steps in data collection, data preparation and data analysis were described thoroughly. Furthermore access to the used database and R scripts for data collection, data transformation and data analysis can be used so the study can be replicated. Furthermore, data is used which is publicly accessible.

Internal validity includes the ability to conclude relations between variables from correlations. This is aimed for by using the appropriate models and testing for the right assumptions. Furthermore it is specifically mentioned what types of variance can be explained with the statistical analyses and what not.

External validity reflects the generalizability of the research. As data is used which includes (nearly) the whole population of countries, this study aims to gain high generalizability. However, the data used includes some biasses and under and over representation. Implications will be further discussed in the discussion chapter.

Construct validity concerns the proper representation of concepts by the chosen indicators. By using indicators or measures that are commonly used or developed by researchers in the field of scientometrics, evolutionary economy or other relevant studies, this study aims to provide reasonable construct validity.

Lastly robustness checks were included in different parts of the analysis in order to find out whether choices in calculations and data transformation had a large impact on the outcomes of the analysis. As will be discussed those resulted in limited deviations of the results, consolidating the robustness of the analyses.

4. Results

The results chapter consists of four parts. First in section 4.1 the constructed publication dataset is presented using descriptive statistics, plots and visualizations, in order to illustrate the distribution of variables and general trends in the data. Next, in section 4.2, first analyses are performed to answer hypotheses 1 and 2, using results from statistical analysis of the publication data set in combination with panel data on economic growth from The World Bank. Then, the following part zooms in on the content of the portfolio and how this is related to different stages in economic development as described in hypothesis 3. The final part presents the results of analyses conducted to explore and test relations between concepts more specifically aimed at certain regions or clusters of countries and their institutional context, in order to test hypothesis 4.

4.1 Descriptive statistics

Publication database
Data retrieved from Scimago JR was used to construct a publication database that could be employed to study trends in publications over scientific fields and areas. The database consists of data on 104.290.863 publications, divided over 21 years (1996-2016), 239 countries, 27 categories (areas of science) and 310 subcategories (fields of science) - for an overview of the areas and fields see Appendix A.

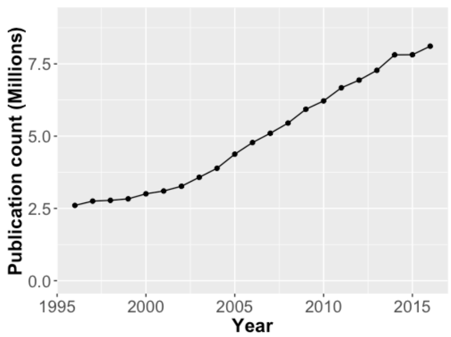


Figure 3. - Worldwide publications per year.

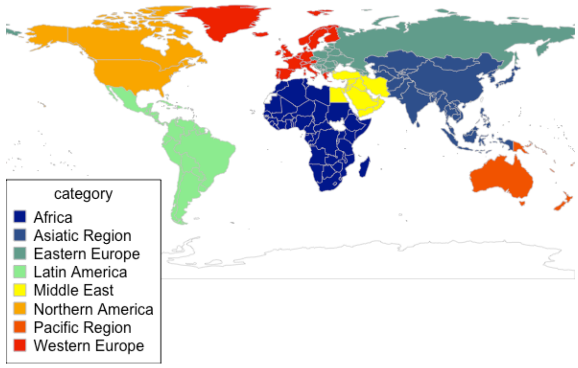


Figure 4. - Overview of the 8 world regions.

The number of publications that are added per year has increased drastically over the 21 year period as can be seen in figure 3. Ranging from 2,60 million in 1996 to 8,11 million in 2016 worldwide, equal to an on average 14,8% increase in growth every year. Furthermore the total amount of publications also varies greatly between world regions as seen in table 4.

World Region	Documents	% of total
Western Europe	32889623	31,5
Northern America	27403488	26,3
Asiatic Region	26274438	25,2
Eastern Europe	5979669	5,7
Middle East	3854233	3,7
Latin America	3442098	3,3
Pacific Region	3016833	2,9
Africa	1430481	1,4

Table 4. - Share of world regions.

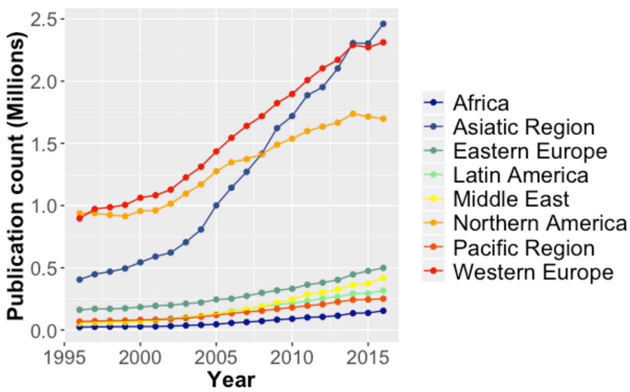


Figure 5. - Increase publications per year over time.

Furthermore the countries are divided in 8 world regions: Africa, North America, Latin America, Pacific Region, Asiatic Region, Middle East, West Europe and East Europe. The specific distribution of countries over the different world regions can be seen in figure 4. The pacific region consists of the countries in Oceania plus small islands in the Pacific Ocean which are not clearly visible on the map. In appendix B a list of countries and the world region they belong to is provided.

When looking at the development of world regions over time, see figure 5, the most important shift is probably the Asiatic region growing increasingly faster than North America which is slowing down and also overtaking Western Europe. These three regions together account for most of the worldwide increase in publication output. Other world regions also increase but remain relatively small compared to the biggest three. However, individual countries from these regions can still account for large shares of publications or show extensive growth patterns.

The dataset further includes data on an aggregate level, for every country-subcategory-year combination, such as average amount of citations and Hirsch-index value. For more insight in the structure of the data see the description of the R scripts in appendix F. Besides absolute numbers of publications the database contains more detailed information on the categories (from here on referred to as research areas) and subcategories (further referred to as research fields) which countries publish in.

Publication and external sources variables

For the main variables of this study, from the publication database and external sources, the number of observations, the mean, standard deviation, median, minimum and maximum values are provided in table 5. Histograms providing more insight in the distribution of variables can be found in Appendix B.

Variable	nr. obs	mean	sd	median	min	max
publication.count	4684	2.23e+04	9.96e+04	400	1	1.50e+06
entropy	4684	6.00e+00	2.00e+00	6	0	8
ubiquity	4684	7.00e+01	1.60e+01	68	14	1.48e+02
nr.of.fields	5019	1.25e+02	108	95	0	308
hoover.specialization	4684	0.57	0.27	0.57	0.08	1
citations	4684	3.96e+05	2.19e+06	5490	0	3.80e+07
H index	4684	6.16e+03	1.12e+04	1350	0	8.88e+04
specialization (% amounts)	4471	73.3	31.1	87.1	0	100
rel.div. (% amounts)	4471	2.93	5.53	1.18	0	100
unrel.div (% amounts)	4471	23.7	29.7	9.96	0	100
GDP	4039	2.72e+11	1.18e+12	1.44e+10	1.23e+7	1.86e+13
logGDP	4039	2.40e+01	2.00e+00	23	16	3.10e+01
Control of Corruption	3468	0	1	0	-2	2
Government Effectiveness	3460	0	1	0	-2	2
Political Stability and Absence of Violence/T	3485	0	1	0	-3	2
Regulatory Quality	3461	0	1	0	-3	2
Rule of Law	3527	0	1	0	-3	2
Voice and Accountability	3521	0	1	0	-2	2

Table 5. - Country level variables from publication data and external sources.

For the external sources especially the high variation and standard deviation in GDP stands out. For this reason the log of the GDP was taken as a new variable for further analyses (logGDP).

The correlation matrix in table 8 (next page) shows how the variables are related. Some remarkably high correlations are found. Some variables such as publication count, amount of citations and H index could be expected to be highly correlated. Other such as ubiquity, Hoover specialization of the portfolio, entropy or the number of fields are correlated as their calculation depends on similar steps.

However, some correlations can already provide insight in the relations between concepts that this study aims to explore. For instance further exploration of the high correlation between logGDP and entropy, specialization rate or unrelated diversification rate may provide interesting results. Also the correlation between some world governance indicators and several knowledge development aspects may be due to the fact that there is a connection between those contextual factors and knowledge development.

Scientific areas and fields

Another way of looking at the data is through the focus on different scientific areas and underlying scientific fields. The amount is not equally distributed over different areas as can be seen in table 6. Medicine, engineering and biochemistry are the largest areas. Dentistry, veterinary and decision sciences are almost a factor 100 smaller.

Scientific Area	areacode	Publications
Medicine	2700	21255946
Engineering	2200	11276860
Biochemistry, Genetics and Molecular Biology	1300	9223245
Physics and Astronomy	3100	7709216
Computer Science	1700	7494590
Materials Science	2500	7005712
Chemistry	1600	4909196
Agricultural and Biological Sciences	1100	4861947
Mathematics	2600	4125488
Earth and Planetary Sciences	1900	3581908
Environmental Science	2300	3318551
Social Sciences	3300	3193607
Chemical Engineering	1500	1971214
Pharmacology, Toxicology and Pharmaceutics	3000	1925319
Immunology and Microbiology	2400	1779000
Energy	2100	1637103
Arts and Humanities	1200	1527762
Neuroscience	2800	1508661
Business, Management and Accounting	1400	1152538
Psychology	3200	1142145
Health Professions	3600	746160
Economics, Econometrics and Finance	2000	677595
Nursing	2900	676580
Multidisciplinary	1000	472364
Decision Sciences	1800	462912
Veterinary	3400	412459
Dentistry	3500	242785

Table 6. - Overview of Scientific areas in order of amount of publications.

When comparing the amount of publications per field in 1996 and 2016, we can see large differences as well. The 10 fields that have increased the most in growth account for an addition in over 1 million publications in 2016, almost 20% of the increase in growth over the whole period. These are presented in table 7. This shows that some fields in medicine and computer science have especially increased in two decades. Further variations of areas and fields over time and space are elaborated on in the next section.

Field	**Area**	**Increase in publications from 1996 to 2016**
Software	Computer Science	72019
Condensed Matter Physics	Physics and Astronomy	75316
Mechanical Engineering	Engineering	82894
Materials Science (miscellaneous)	Materials Science	96837
Chemistry (miscellaneous)	Chemistry	100165
Oncology	Medicine	101324
Computer Networks and Communications	Computer Science	110218
Computer Science Applications	Computer Science	112068
Electrical and Electronic Engineering	Engineering	153507
Medicine (miscellaneous)	Medicine	174236

Table 7. - Top 10 fields with the highest increase in growth - 1996 to 2016.

Variables	publica	entropy	ubiquity	nr.of.fie	hoover.:	citation	H index	speciali	rel.div. (	unrel.di	GDP	logGDP	Control	Govern	Politica	Regulat	Rule of	Voice a
publication.count	1	0.22	-0.23	0.34	-0.32	0.81	0.75	0.19	-0.12	-0.18	0.97	0.47	0.23	0.27	0.08	0.24	0.24	0.16
entropy	0.22	1	-0.51	0.85	-0.91	0.18	0.49	0.83	-0.22	-0.83	0.21	0.81	0.27	0.39	-0.02	0.4	0.24	0.11
ubiquity	-0.23	-0.51	1	-0.55	0.59	-0.23	-0.47	-0.42	0.11	0.42	-0.23	-0.54	-0.37	-0.48	-0.15	-0.47	-0.38	-0.2
nr.of.fields	0.34	0.85	-0.55	1	-0.97	0.27	0.7	0.79	-0.35	-0.77	0.31	0.89	0.38	0.5	0.06	0.49	0.37	0.22
hoover.specialization	-0.32	-0.91	0.59	-0.97	1	-0.27	-0.67	-0.82	0.3	0.8	-0.3	-0.88	-0.38	-0.5	-0.06	-0.49	-0.37	-0.21
citations	0.81	0.18	-0.23	0.27	-0.27	1	0.72	0.16	-0.09	-0.14	0.81	0.39	0.25	0.27	0.1	0.24	0.25	0.19
H index	0.75	0.49	-0.47	0.7	-0.67	0.72	1	0.44	-0.25	-0.41	0.72	0.75	0.52	0.57	0.22	0.53	0.51	0.39
specialization (% amounts)	0.19	0.83	-0.42	0.79	-0.82	0.16	0.44	1	-0.33	-0.98	0.18	0.74	0.17	0.3	-0.07	0.32	0.16	0.03
rel.div. (% amounts)	-0.12	-0.22	0.11	-0.35	0.3	-0.09	-0.25	-0.33	1	0.16	-0.13	-0.36	-0.19	-0.25	-0.07	-0.24	-0.21	-0.17
unrel.div (% amounts)	-0.18	-0.83	0.42	-0.77	0.8	-0.14	-0.41	-0.98	0.16	1	-0.17	-0.72	-0.15	-0.28	0.08	-0.3	-0.13	0
GDP	0.97	0.21	-0.23	0.31	-0.3	0.81	0.72	0.18	-0.13	-0.17	1	0.45	0.21	0.24	0.06	0.22	0.22	0.14
logGDP	0.47	0.81	-0.54	0.89	-0.88	0.39	0.75	0.74	-0.36	-0.72	0.45	1	0.29	0.44	-0.04	0.42	0.26	0.11
Control of Corruption	0.23	0.27	-0.37	0.38	-0.38	0.25	0.52	0.17	-0.19	-0.15	0.21	0.29	1	0.92	0.74	0.86	0.94	0.76
Government Effectiveness	0.27	0.39	-0.48	0.5	-0.5	0.27	0.57	0.3	-0.25	-0.28	0.24	0.44	0.92	1	0.7	0.93	0.93	0.75
Political Stability and Absence	0.08	-0.02	-0.15	0.06	-0.06	0.1	0.22	-0.07	-0.07	0.08	0.06	-0.04	0.74	0.7	1	0.65	0.78	0.68
Regulatory Quality	0.24	0.4	-0.47	0.49	-0.49	0.24	0.53	0.32	-0.24	-0.3	0.22	0.42	0.86	0.93	0.65	1	0.9	0.77
Rule of Law	0.24	0.24	-0.38	0.37	-0.37	0.25	0.51	0.16	-0.21	-0.13	0.22	0.26	0.94	0.93	0.78	0.9	1	0.82
Voice and Accountability	0.16	0.11	-0.2	0.22	-0.21	0.19	0.39	0.03	-0.17	0	0.14	0.11	0.76	0.75	0.68	0.77	0.82	1

Table 8. - Correlation matrix.

4.2 Analyses

Knowledge space

Hypotheses 1 and 2 focus on characteristics of the knowledge space. The distribution or division of countries based on their portfolio of scientific publications, and their development over time reflected by their movement in this space. These were tested as follows.

The measure of entropy was used to provide insight in the changes of knowledge of different countries over time and over different levels of GDP. The value of entropy reflects the complexity of the knowledge published by countries, by taking into account portfolio aspects as the variety of publications in different fields and the amount of publications in these fields.

When this is visualized in a plot (figure 6), a general trend can be seen, it seems that the entropy of countries' research portfolios increases over time, and when GDP is higher. Individual trajectories of countries can also be visualized as seen in the right plot, the US (above) and Canada (below) both show a general trend of increase of entropy and GDP over time.

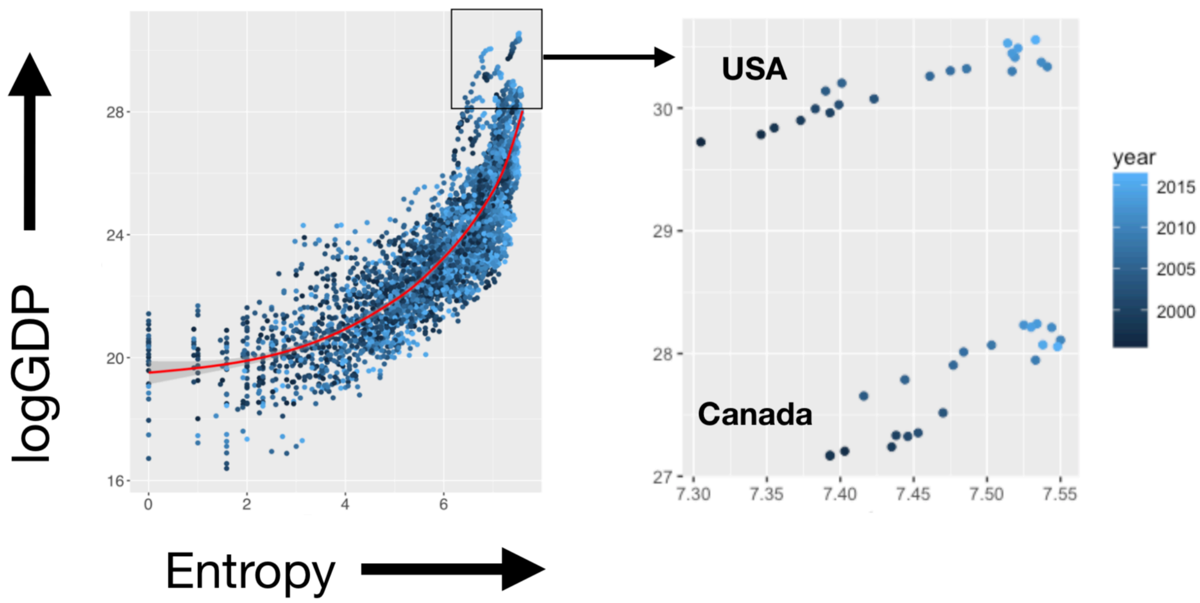


Figure 6. - Entropy vs. logGDP for all countries (left), and for US and Canada (right), from 1996-2016.

The majority of countries has such a development trajectory in the same direction, both increasing in GDP and in entropy. This results in the fact that the mean entropy of all countries increases from 5.14 (1996) to 5.93 (2016) and the logGDP from 23.00 (≈ 1.67 billion USD) to 24.16 (≈ 404.13 billion USD). When logGDP is plotted against entropy for individual years a similar trend is seen each year as in figure 7. This is shown below for 1996, 2006 and 2016.

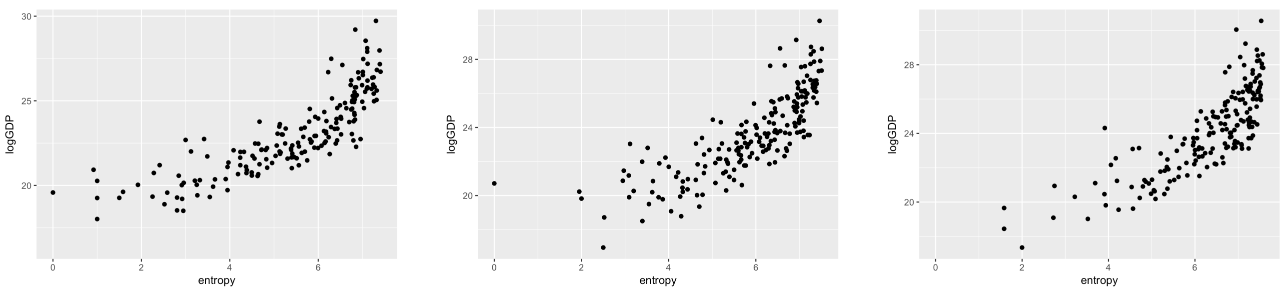


Figure 7. - Entropy vs. logGDP in 1996, 2006 and 2016.

However, there is still a broad range of possible positions on the GDP axis for the same entropy values, and similarly a broad range of possible positions on the entropy axis for countries with the same GDP. This will be further explored by looking at the influence of other variables on both GDP and entropy.

Plots can also be used to illustrate where different world regions are located in the knowledge space. Figure 8. Shows that different regions have a distinctive position in the knowledge space. The plots are based on all publications in a world region over the whole period from 1996 to 2016.

This shows that countries in Eastern Europe, the Middle East, Western Europe and Northern America have quite concentrated positions in terms of entropy and GDP. The Asiatic region, Africa and Latin America are more spread over the whole spectrum of entropy and GDP. For the Pacific Region it can be seen that Australia and New Zealand are positioned in the top right with high knowledge complexity and GDP, while the other countries are positioned in the bottom left, with lower entropy and GDP.

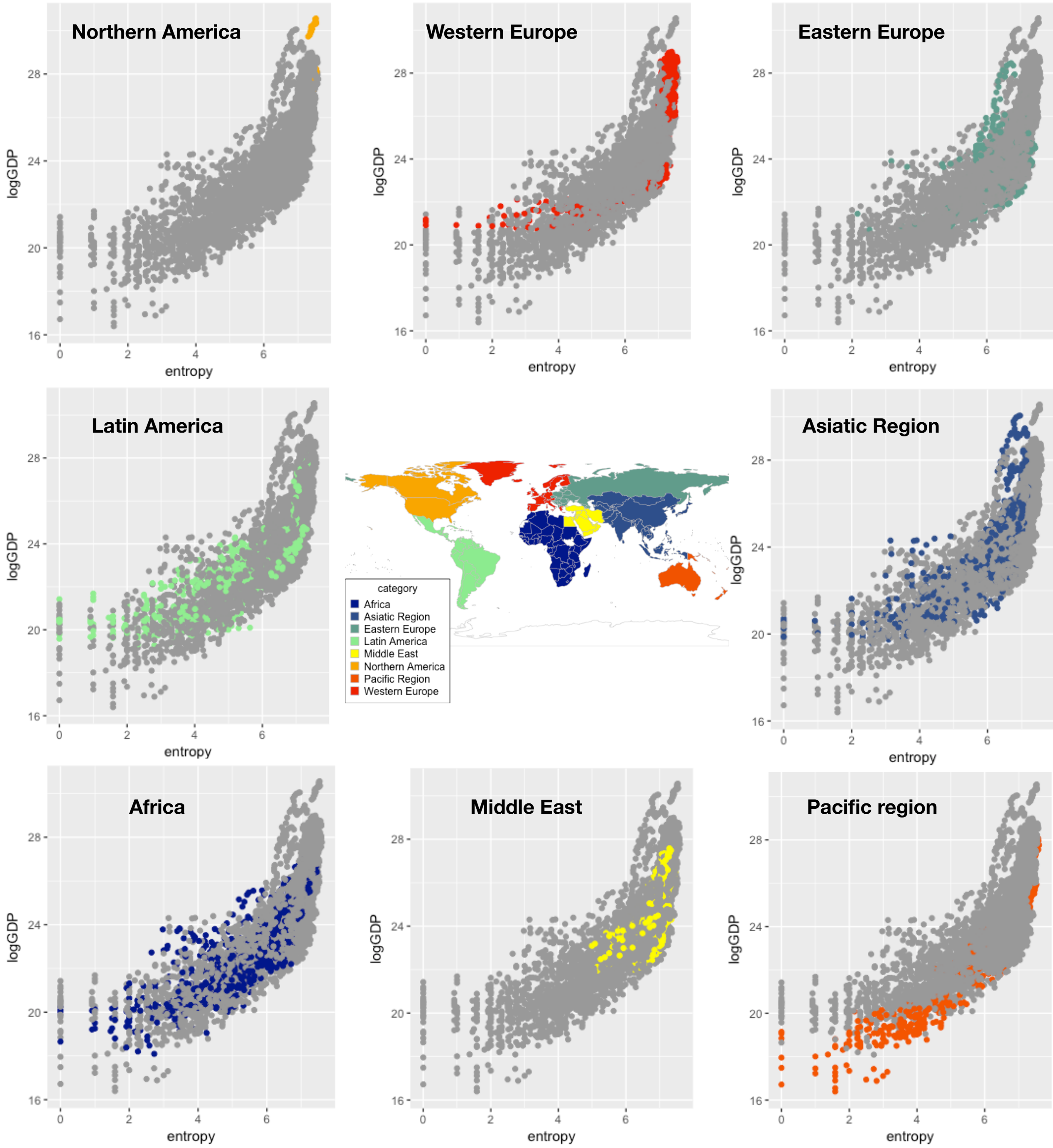


Figure 8. - Overview of the 8 world regions and their position in terms of GDP and entropy.

Statistical analysis

By looking at the data from the perspective of entropy and GDP, and the development trajectory that seems to exist, the question rises; what determines a countries' position in the knowledge space, and its development over time? By performing regression analyses we can see how different portfolio characteristics are related to economic development in terms of GDP and knowledge development reflected by entropy.

Entropy and GDP

First the correlation between entropy and logGDP, as expected from the trend that can be seen in the plots above, is tested. A panel linear model regression analysis was used with included fixed effects for individual countries (201) and years (1996-2016). This means that the regression isolated country context that may influence the dependent variable, and controls for the aggregate variations that exist over time.

Table 9 provides a summary on the results of this model. More elaborated output of all statistical regressions is provided in appendix E, as well as an explanation on the performed assumption tests to choose the right model in appendix D. Indeed a positive significant correlation is found, meaning that within country variation in GDP over time can be explained by variation in entropy. The R^2 value of 0.79 also indicates that the model is able to explain a high percentage of the variation in logGDP. This is partly with entropy and partly with the fixed effects on country and year level.

Model	IV	Effect (+ significance)	DV	R^2	N
1	Entropy	0.0218543 (0.0144449) *	logGDP	0.79321	3998

Table 9. - PLM regression analysis entropy vs. logGDP.

The found estimate means that for one unit increase in entropy of the knowledge portfolio, logGDP is expected to increase by 0.0218543 units, or corrected for the logarithmic variable of GDP; a 1-unit increase in entropy multiplies the expected value of GDP by 1.022095. While this may seem like a small effect, a yearly increase of the total GDP on national level of a few percentages is a large effect. Further exploration of the estimated fixed effects and coefficients of specific countries can show in what countries the correlation between entropy and logGDP plays a more important role and in what countries this is smaller.

Process of specialization, related and unrelated diversification

The extent to which countries specialize or diversify into related or unrelated fields can be compared to their entropy and logGDP, reflecting their knowledge and economic development. The regression analysis shows (see table 10 and 11) that within and across countries specialization seems to have a significant positive effect on both entropy and logGDP. Not surprisingly the effect of diversification is the opposite, as a higher percentage specialization automatically means a lower percentage diversification in the portfolio and vice versa.

Model	IV	Effect (+ significance)	DV	R^2	N
2	Specialization	**0.00164799** (4.618e-06) ***	logGDP	0.79494	3814
3	Related diversification	**-0.00430309** (1.625e-05) ***	logGDP	0.7948	3814
4	Unrelated diversification	**-0.00112089** (0.002148) **	logGDP	0.79428	3814
5	Specialization	**0.00137349** (0.0001898) ***	logGDP	0.7956	3814
	Related diversification	**-0.00346573** (0.0006861) ***			
6	Specialization	**0.0048392** (1.542e-06) ***	logGDP	0.7956	3814
	Unrelated diversification	**0.0034657** (0.0006861) ***			

Table 10. - PLM regression analysis results summary - DV logGDP.

Estimates for model 2, 3 and 4 can be interpreted as follows; a one percent increase in specialization is expected to increase entropy by 0.22%. For model 7, 8 and 9 the estimates are interpreted similar to those in model 1. Differences in the amount of observations depends on the amount of countries that data was available for, and on the way that measures were calculated. For instance specialization loses one year of data as the measure is determined by comparing two consecutive years.

Model	IV	Effect (+ significance)	DV	R^2	N
7	Specialization	0.00220215 (0.0002761) ***	entropy	0.37258	3814
8	Related diversification	0.0074665 (2.119e-06) ***	entropy	0.37396	3814
9	Unrelated diversification	-0.00330396 (4.690e-08) ***	entropy	0.37505	3814
10	Specialization	0.00285741 (3.392e-06) ***	entropy	0.37716	3814
	Related diversification	0.00889194 (2.807e-08) ***			
11	Specialization	-0.0060345 (0.0001621) ***	entropy	0.37716	3814
	Unrelated diversification	-0.0088919 (2.807e-08) ***			

Table 11. - PLM regression analysis results summary - DV entropy

What is interesting is the fact that for the same specialization rate, unrelated diversification has a positive effect on logGDP but a negative effect on entropy, while related diversification has a negative effect on entropy and a positive effect on logGDP. While the first results suggest more developed countries specialize more and diversify less than less developed countries, the effect of related versus unrelated diversification seems to be more complex. However, the positive correlation of unrelated diversification and logGDP is in line with the innovation theories stating that moving into unrelated fields increases comparative advantage, enabling economic growth.

To look further than within-country variation over time, trends over all observations can be explored. In order to see how related and unrelated diversification and specialization vary over different stages in economic development, classes can be made to divide all observations into 10 equal size groups of different ranges of logGDP values. When this is done the mean for each class

Class	Specialization (%)	Related diversification (%)	Unrelated diversification (%)
1	29.47	4.74	65.79
2	48.64	5.13	46.23
3	59.46	4.89	35.65
4	72.44	4.12	23.44
5	82.27	3.18	14.55
6	85.70	3.01	11.29
7	93.28	1.84	4.89
8	97.52	0.95	1.53
9	99.69	0.21	0.10
10	99.96	0.03	0.01

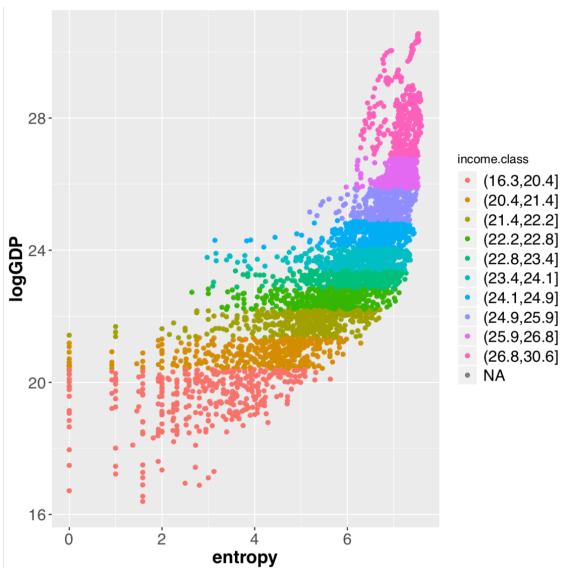


Table 12. - Specialization and diversification over different stages in development.

is determined as shown below in table 12. This classification is used multiple times over the data analysis. In order to test the robustness of this classification different amounts of classes have been used. This showed no significant difference in results.

Table 12 shows that in the lowest income classes the biggest share of publications added to a portfolio are in unrelated diversification. However, when GDP increases specialization takes up a larger part of knowledge development and both related and unrelated diversification become smaller and smaller until there is only a fraction left.

This global trend seems to be contradictory to the results of table 10 and 11 showing some negative correlations of diversification with entropy and within country variation in GDP over time. However, as shown by the regression analysis when unrelated diversification is increased in a country GDP increases as well. This may mean that even though countries with a larger portfolio diversify less and less, increasing unrelated diversification can be profitable.

Furthermore the development trajectory of knowledge development may require countries with a small knowledge portfolio and lower economic wealth to first develop a broader set of knowledge by entering new fields through related and unrelated diversification in order to be able to develop further and to be able to later specialize in certain fields.

Other portfolio characteristics

In order to further explore how different aspects of the portfolio are related to knowledge complexity and economic growth other characteristics are also analyzed. The following models explore how other variables based on the publication database may correlate with entropy and GDP. As shown in table 13 ubiquity and publication count (total amount of publications) of a countries' portfolio seems to have a negative effect on both GDP and entropy. The latter is a surprising effect as one might expect that growth in publications would have a positive effect in both knowledge and economic development.

Model	IV	Effect (+ significance)	DV	R^2	N
12	Ubiquity	-0.01202397 (< 2.2e-16) ***	Entropy	0.39889	4684
13	Nr. of fields	0.00597758 (< 2.2e-16) ***	Entropy	0.40828	4684
14	Hoover coefficient	-5.8940378 (< 2.2e-16) ***	Entropy	0.58287	4684
15	Publication count	-1.8525e-06 (< 2.2e-16) ***	Entropy	0.38733	4684
16	Citation count	2.8528e-08 (0.003643) **	Entropy	0.37872	4684
17	H index	2.0006e-04 (3.703e-08) ***	Entropy	0.38178	4684
18	Ubiquity	-0.00339939 (3.167e-07) ***	logGDP	0.79431	3998
19	Nr. of fields	0.00382575 (< 2.2e-16) ***	logGDP	0.80894	4039
20	Hoover coefficient	-0.906832 (< 2.2e-16) ***	logGDP	0.79928	3998
21	Publication count	-5.7352e-08 (0.6043188)	logGDP	0.7929	3998
22	Citation count	3.5573e-08 (2.477e-13) ***	logGDP	0.7958	3998
23	H index	2.1082e-04 (< 2.2e-16) ***	logGDP	0.79963	3998

Table 13. - other portfolio characteristics vs. entropy and logGDP.

The number of fields that are present in a portfolio, similar to the diversity or variety within a portfolio, has a positive effect on both, suggesting that diversity is more important than size. This is reinforced by the remarkably high values of the estimates of the correlation between the Hoover specialization coefficient and entropy and logGDP. This captures the concentration in a small amount of fields. The effect indicates that a more specialized portfolio has a lower entropy, which could also be due to the fact that lower variety inherently means a lower entropy due to the way it is calculated. However, the strong correlation with logGDP indicates that it is also negatively correlated to economic growth. This suggests that broadening the portfolio through diversification would be preferable to specializing.

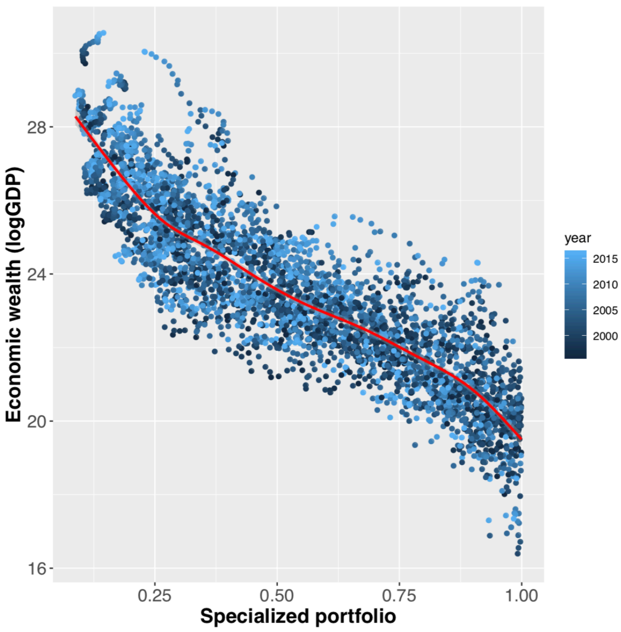


Figure 9. - Specialized portfolio vs. logGDP.

When the specialization coefficient is plotted against GDP, see figure 9, a clear trend is also visible in line with the found correlation; portfolios with a higher GDP have a lower specialization coefficient and vice-versa. Zooming in on some specific trajectories of observations it also seems that specific countries move to a higher GDP and a lower specialization coefficient at the same time.

Lastly, the citation count and the average H-index of a countries' portfolio have a positive effect on both entropy and GDP. This could mean that countries which increase their portfolios quality in terms of a higher citation count and H index produce more complex and valuable knowledge as this also reflects positively in their economic growth.

The fact that the correlation with publication count is negative would suggest that it is not absolute portfolio size, but variety and quality of publications that matters.

Conclusions on hypotheses 1 and 2

The first results of the analysis seems to be in line with the expectations of hypothesis 1 and 2. It suggests that there is indeed a variation in knowledge portfolio's complexity over countries. Trajectories of development from less complex knowledge to more complex knowledge over time can be discerned. The analysis also shows that developments of countries over time in terms of economic growth (increases in GDP) correspond to countries development in terms of knowledge complexity (reflected by entropy). Also the correlation of ubiquity with GDP and entropy shows that when a portfolio develops knowledge that is more common, and it can be assumed less valuable or complex, this has a negative effect on the development of that country over time.

Furthermore, other variables such as rate of specialization, the number of fields and publication and citation count can be used to further specify and possibly explain how countries develop scientific knowledge, and its position in terms of entropy and GDP. Insight in the process of specialization vs. related and unrelated diversification also seems to help explain variance over time. The division of data into different income classes illustrates variation of diversification and specialization in the portfolio along the development trajectory.

However, the expectations as described in the theory chapter and first hypothesis on the role of unrelated and related diversification are not entirely met by the analysis results. It is not evident why unrelated diversification would have a negative effect on GDP, while controlled for specialization this effect would become positive. Also the negative correlation of unrelated diversification, also when controlled for specialization, with entropy is surprising, as one might expect that a portfolio would become more complex as unrelated diversification could lead to a portfolio containing a higher variety of different types of knowledge.

Further analysis of different types of areas and fields of scientific knowledge and their distribution over time and space may provide more clues.

Research fields & areas

While the panel data model regression analysis is a good way to explore whether expected relations between portfolio characteristics and other country characteristics exist in the data, they only explain within-country variance over time and they do not reflect the actual content of a portfolio. In order to explore the variation between countries, a different perspective on the data can be taken. This is done by focusing on the different fields of scientific publications, their respective growth and in which countries they grow most or take up the largest share of publications.

As stated earlier, every scientific field has grown over the past decades, as the total output of scientific publications has increased from 2,60 million in 1996 to 8,11 million in 2016 worldwide. Table 6 and 7 already illustrated the distribution of areas over the whole period, and the 10 fields with the biggest increase from 1996 to 2016. When taking a look at the developments of different fields over time, as reflected by worldwide publications each year, it can be seen that in general all fields grow increasingly over time - see figure 10 (next page). Furthermore in some fields much more publications are added each year. *Medicine* takes the lead in terms of yearly publication output, followed by *engineering* and *computer science*. Whats remarkable is the increase of *engineering* over *biochemistry, generics and molecular science* and the field of *computer sciences* overtaking five other fields in the last 20 years.

Respective positions of areas in an average portfolio can also be compared, using the percentage an area takes in in the global publication count each year, as illustrated in figure 11. There it can be seen that for instance *energy* and *environmental science* have increased in relative share over time, as well as *social sciences. Physics, biochemistry* and *immunology* have decreased in share over time. While the share may not reflect whether a scientific discipline has become bigger in absolute scientific output, it may help to evaluate the role of a specific discipline in the global science system in general and in specific country portfolios in particular. Also a sustained increase in share may be an indicator that a field is emerging.

Growth factor

In order to compare areas and fields in terms of growth over time, a growth factor can be determined. When correcting for the average growth per year, different scientific areas and fields can be compared on their relative growth (above or below average) and the relative increase in growth from 1996 to 2016 can be determined. In order to take into account not only the start and end values, but every year in between, a linear model was used to determine the average increase over the whole period. This was consequently used to calculate the growth factor, or more correctly the relative-increase-in-growth factor, per scientific area and field.

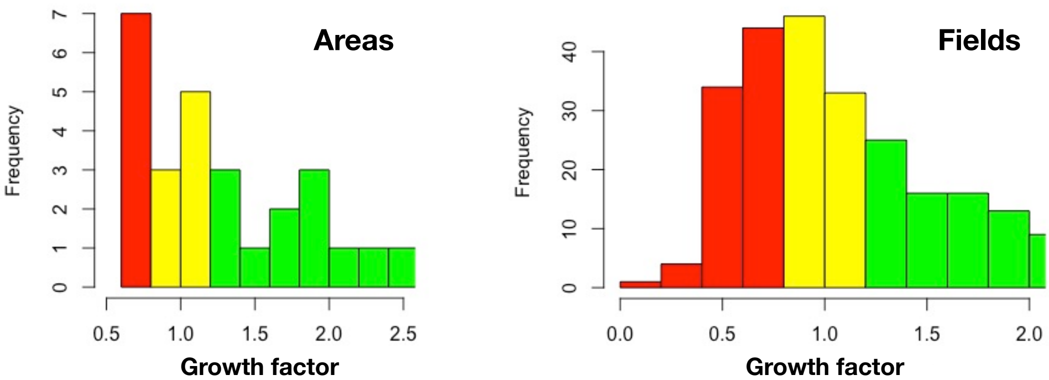


Figure 12. - Histograms of areas and their growth factor, and fields and their growth factor.

As shown in the histograms in figure 12. These growth factors, reflecting relative increase in growth, of the 27 research areas, and the underlying 310 research fields is distributed around 1, meaning average increase in growth. The range of growth factors was used to classify areas and fields as “decline”: 0-0.8, a decline in growth, “stable”: 0.8-1.2, a relatively stable increase in growth or a slightly above or below average increase, or “growth”: above 1.2, a relatively fast increase in growth of fields and areas as compared to others.

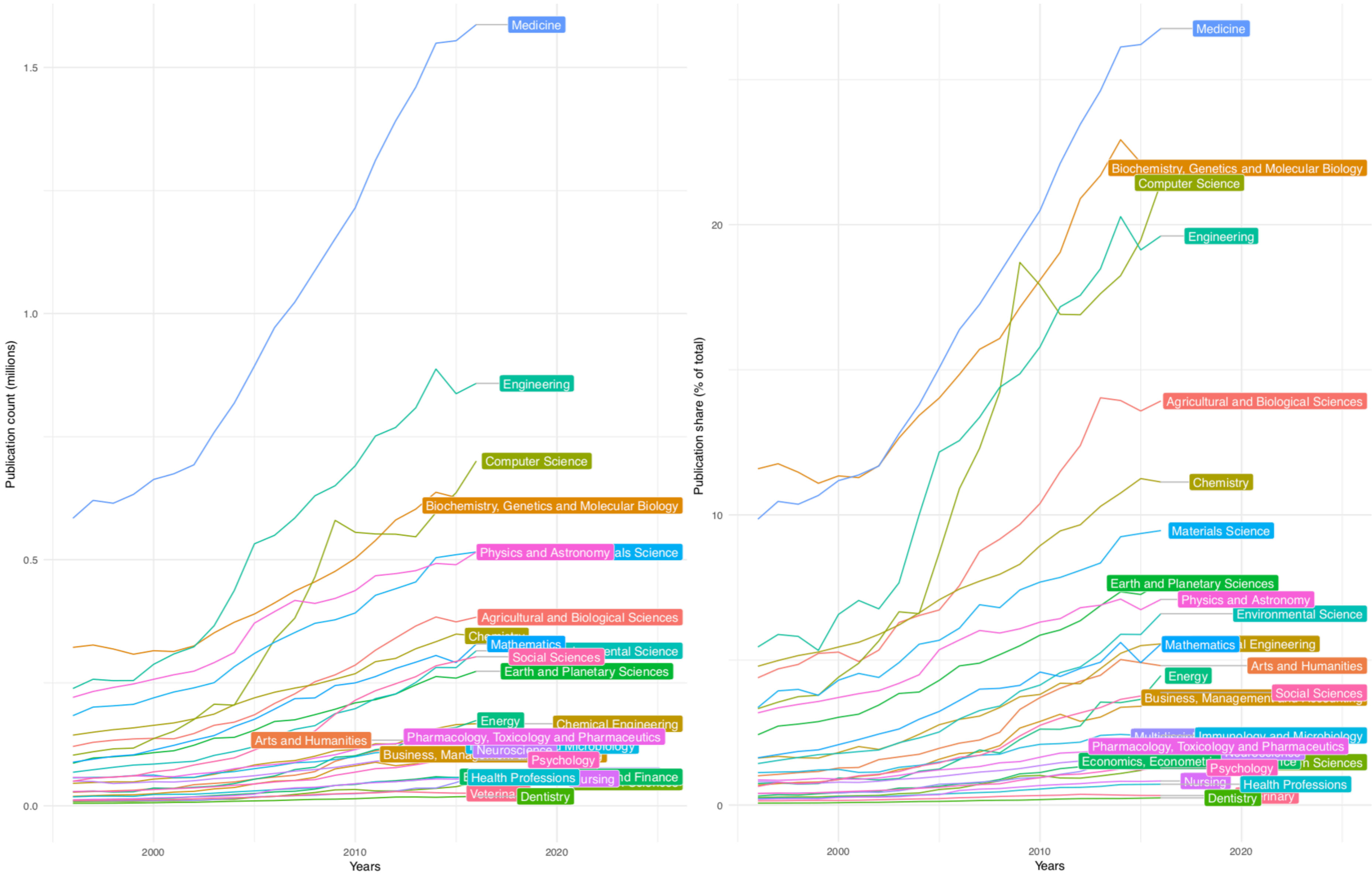


Figure 10 & 11. - Publication count and share of total for each area over time.

The highest areas in the “growth” category include for instance Computer Science (2.52 growth factor), Energy Science (2.22) and Social Sciences (2.08). “Stable” areas are Agricultural and

Biological Sciences (1.05), Psychology (1.04) and Health Professions (1.00). Areas in “decline” are Chemistry (0.71), Physics and Astronomy (0.69) and Pharmacology, Toxicology and Pharmaceutics (0.80).

Note that fields that are part of a declining scientific area are not necessarily in decline as well, they can be stable or grow while others cause the average decline of the area. Research fields with high growth rates are for instance some fields in the area of Computer Sciences: Human-Computer interaction (10.3), Artificial Intelligence (4.1) or in the area of Energy: Renewable Energy, Sustainability and the Environment (4.6).

Using these three categories of relative growth of fields, a countries’ portfolio can be evaluated on the amount of publications in each category it contains. This reflects the way different countries are fit to enable growth of fast growing, emerging fields, or whether they produce more of their knowledge in more stable and declining fields. When looking at the whole period of 1996 to 2016, countries with a high percentage of publications in “growth” are for instance Hong Kong (48%) and the United Arabian Emirates (50%). Countries with a low percentage are for instance Russia (21%) and Germany (25%).

Zooming in on the specific growth of areas in the portfolio of Hong Kong over time shows that in recent years Computer Science and Engineering where the areas that had the highest publication count in Hong Kong. Furthermore, while Hong Kong is a small countries, in some areas such as D*ecision sciences*, B*usiness, management and accounting* and *Economics, econometrics and finance*, which are areas that grow above average, Hong Kong contributes up to one or in some years two percent of the global publication output. On the contrary, Russia’s areas with the highest publication output are more traditional and below average growing fields such as *Physics and astronomy* and *Materials science.*

This shows that the growth factor of a portfolio can provide insight in what type of fields countries specialize in, and whether this seems to be a good sign or not. In case of Hong Kong, either reflecting that they are good at facilitating knowledge development of emerging fields, or that they made a deliberate choice to invest a lot in emerging fields, this seems to be positive. In Russia the science system may be bound to path dependence as it is specialized in more traditional fields and it may be harder to switch to other, faster growing, emerging scientific disciplines.

Average growth factor.
Another way to use these factors is to calculate the average growth factor of a portfolio by taking the weighted average of the publications it contains. This results in the following map - figure 13. - showing the average growth factor per country. This shows for instance that China and Australia seem to be two of the best suited countries enabling emerging fields to grow. While on the other hand Russia, as discussed before, and Argentina on average contain fields that grow below average.

However, it should be noted that this factor represents the weighted average growth rate of all publications in a portfolio, and does not account for the fact that the absolute amount of publications in some areas or fields with a higher growth factor can be bigger.

Stages in development
In order to take a explore how the distribution of different classes ‘decline’, ‘stable’ and ‘growth’ are distributed over different stages of development, the division of portfolios in 10 income classes is used again. Table 14 shows the classes with their respective percentages of fields in the different growth classes, and the average growth factor.

This shows that over the development trajectory (as visualized in plots of entropy vs. logGDP) the percentage of declining fields is the highest for countries with the highest income. The percentage of stable fields is lower for countries with a higher income. The high growth fields percentage, and the average growth factor is the highest in the middle income classes and class 9.

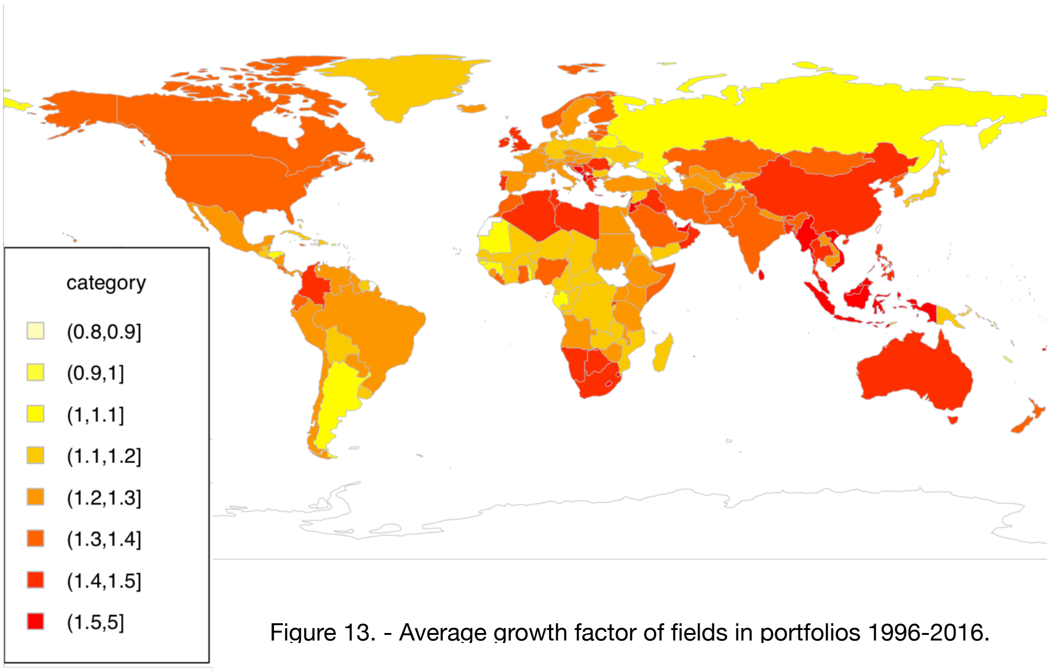


Figure 13. - Average growth factor of fields in portfolios 1996-2016.

While the percentages are not that far apart, there seem to be some differences over the variety of countries across economic development. It could be interpreted that high income countries are traditionally more active in science and therefore still have a relatively large share in traditional disciplines that are not growing that fast anymore. The higher average growth factor of middle income classes could be due to the fact that those are countries that have been successfully developing into higher classes over the last decades by investing in emerging fields of science and technology.

	Entropy class	decline (%)	stable (%)	growth (%)	Average growth factor
1	(16.3,20.4]	27.58	39.59	32.83	1.28
2	(20.4,21.4]	29.13	42.3	28.57	1.22
3	(21.4,22.2]	31.9	39.83	28.27	1.21
4	(22.2,22.8]	30.25	40.06	29.68	1.26
5	(22.8,23.4]	29.04	40.16	30.8	1.34
6	(23.4,24.1]	30.14	39.02	30.84	1.32
7	(24.1,24.9]	30.75	37.89	31.36	1.33
8	(24.9,25.9]	32.56	37.83	29.61	1.28
9	(25.9,26.8]	31.44	35.99	32.57	1.37
10	(26.8,30.6]	35.02	36.44	28.54	1.29

Table 14. - Share of fields in different growth categories and average growth factor for different income classes.

Role of different fields in the development trajectory

As presented before, countries seem to have differing rates of specialization and related or unrelated diversification in their knowledge development, depending on their position in the development trajectory. Also the regressions show that specialization and diversification correlate significantly with GDP and knowledge complexity. It can also be explored how this differs per scientific field. This can be determined by looking how much of the publications in a field are produced as part of specialization or related or unrelated diversification. This results in an overview, of which the largest 10 fields in terms of each category are presented in table 15, 16 and 17.

Field	Area	Specialization (%)	Related diversification (%)	Unrelated diversification
Medicine (miscellaneous)	Medicine	99.99	0	0.01
Electrical and Electronic Engineering	Engineering	99.98	0	0.01
Condensed Matter Physics	Physics and Astronomy	99.98	0.01	0.01
Electronic, Optical and Magnetic Materials	Materials Science	99.97	0.02	0.01
Chemistry (miscellaneous)	Chemistry	99.97	0.02	0.01
Mechanical Engineering	Engineering	99.97	0.02	0.01
Computer Science Applications	Computer Science	99.97	0.01	0.02
Physics and Astronomy (miscellaneous)	Physics and Astronomy	99.97	0.02	0.02
Biochemistry	Biochemistry, Genetics and Molecular Biology	99.97	0	0.03
Materials Chemistry	Materials Science	99.97	0.03	0.01

Table 15. - top ten fields in specialization.

Field	Area	Specialization (%)	Related diversification (%)	Unrelated diversification (%)
Medical Assisting and Transcription	Health Professions	69.65	22.77	7.59
Dental Hygiene	Dentistry	75.43	16.13	8.44
Respiratory Care	Health Professions	18.27	14.5	67.23
Reviews and References (medical)	Medicine	86.09	11.02	2.89
Medical Terminology	Health Professions	68.55	10.48	20.97
Emergency Medical Services	Health Professions	84.46	10.02	5.52
Drug Guides	Medicine	90.95	8.03	1.02
Care Planning	Nursing	92.23	6.75	1.03
Museology	Arts and Humanities	93.17	5.02	1.81
Research and Theory	Nursing	94.03	4.55	1.42

Table 16. - top ten fields in related diversification.

This shows that for some fields such as in table 15 it may be harder to start publishing in this field for a country that has not already published in this field before. For other fields as in table 16 it is relatively easier to enter when a country already published in related fields. For again other fields, such as in table 16, it is the easiest to enter without any publications in either the same area or field before.

Also on average it can be seen that medical research areas, such as *medicine, nursing, dentistry, veterinary* and *health professions*, have fields with on average the highest rates of related and unrelated diversification. Natural sciences, such as *materials science, chemistry, physics*, *agriculture and biological sciences, biochemistry and computer science,* have a lot of fields with on average the a very high percentage in specializations.

Field	Area	Specialization (%)	Related diversification (%)	Unrelated diversification (%)
Respiratory Care	Health Professions	18.27	14.5	67.23
Medical Terminology	Health Professions	68.55	10.48	20.97
Dental Assisting	Dentistry	85.26	2.11	12.63
Dental Hygiene	Dentistry	75.43	16.13	8.44
Medical Assisting and Transcription	Health Professions	69.65	22.77	7.59
Review and Exam Preparation	Nursing	90.12	2.73	7.15
Emergency Medical Services	Health Professions	84.46	10.02	5.52
Nurse Assisting	Nursing	93.34	1.46	5.2
Reviews and References (medical)	Medicine	86.09	11.02	2.89
Respiratory Care	Health Professions	18.27	14.5	67.23

Table 17. - top ten fields in unrelated diversification.

As mentioned before different robustness checks were performed in order to test the reliability of the data and the robustness of the methods to different choices. The results are more or less the same for different methods of classifying fields as related or unrelated. When different thresholds are used, or different association methods are used the results are similar.

When the existing area and field classification of SJR is used to determine relatedness, different specific fields have the highest shares in specialization, related and unrelated diversification. However, still the same trend can be seen that medical fields and areas have higher shares of diversification and areas and fields in natural sciences higher specialization rates.

Portfolio composition

Looking beyond the growth factor of fields as a reflection of the content of a portfolio we can look at specific content of a countries' scientific portfolios too. This is done for the different world regions and different stages in economic development using the income class categorization.

World regions

As found earlier, there is a large difference between world regions in the amount of publications they produce, and in the position they hold in terms of economic development and knowledge complexity. However, the publication data enables to look further into differences in types of publications too. Figure 14 illustrates the average share, that areas take up in a countries' portfolio, of different world regions.

The largest part of the portfolio in any region is taken up with publications in the fields of medicine, biochemistry, engineering, physics, computer science, material science, agriculture, chemistry, mathematics, earth science, social sciences and environmental science. This could be due to the fact that those disciplines have a larger global scientific output in general, or that they are overrepresented in the database. However, the differences between world regions can also tell something about the ability or willingness of countries to produce certain types of knowledge.

Remarkable differences lie in the 'drop' in medicine in Asia and Eastern Europe, where engineering, physics, computer science and material science take up larger parts of the portfolio. This may be because countries in those emerging regions develop in those emerging fields. Furthermore the share of Agricultural and Biological science is relatively high in Africa and Latin America and lower in other world regions. Which may be attributed to the fact that many countries in those continents still rely for a larger part on primary industries. Lastly, the share of Biochemistry is remarkably high in Northern America, which may be related to the fact that there is large Pharmaceutical and Biochemical activity clustered in the US and Canada.

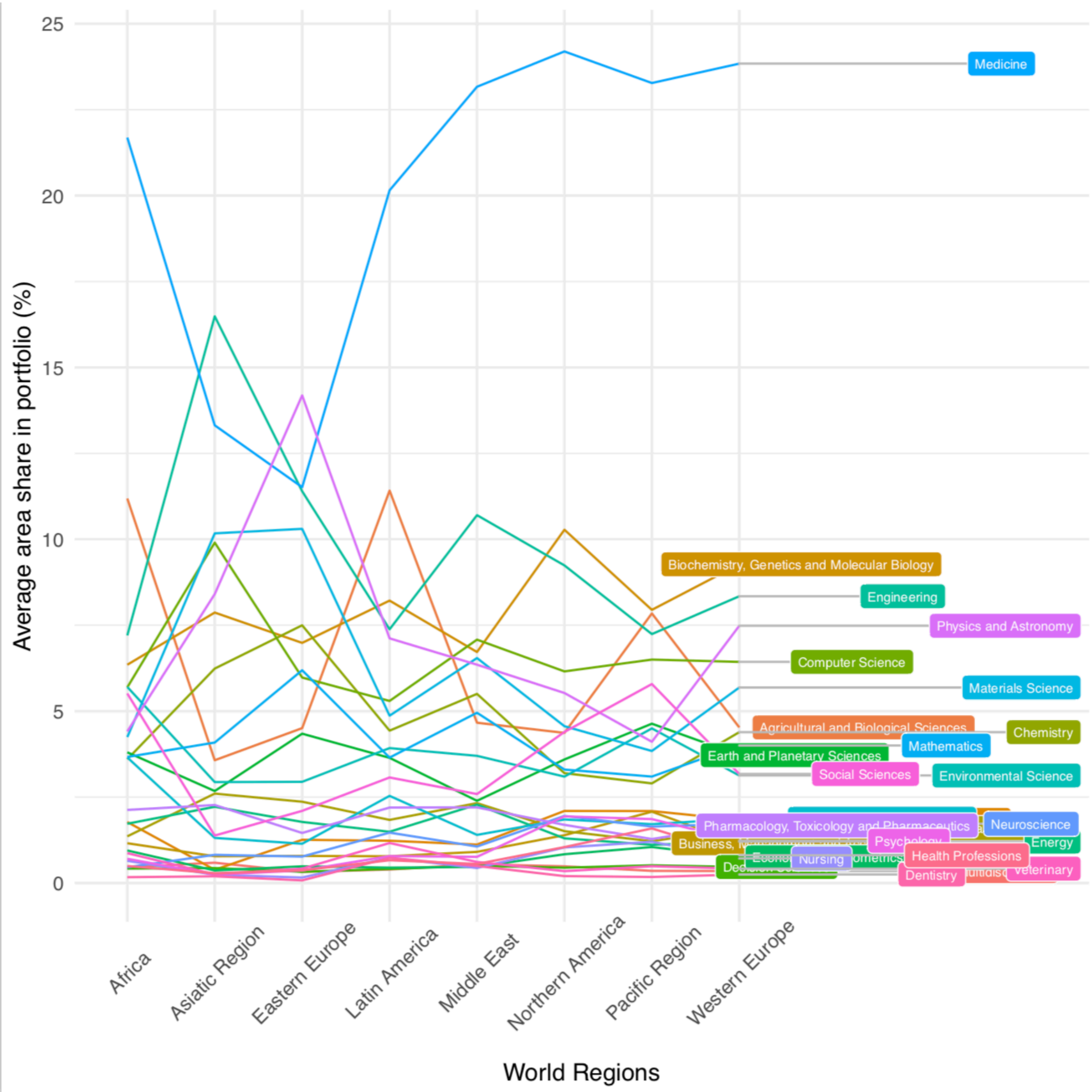


Figure 14. Share of areas in portfolio per world region.

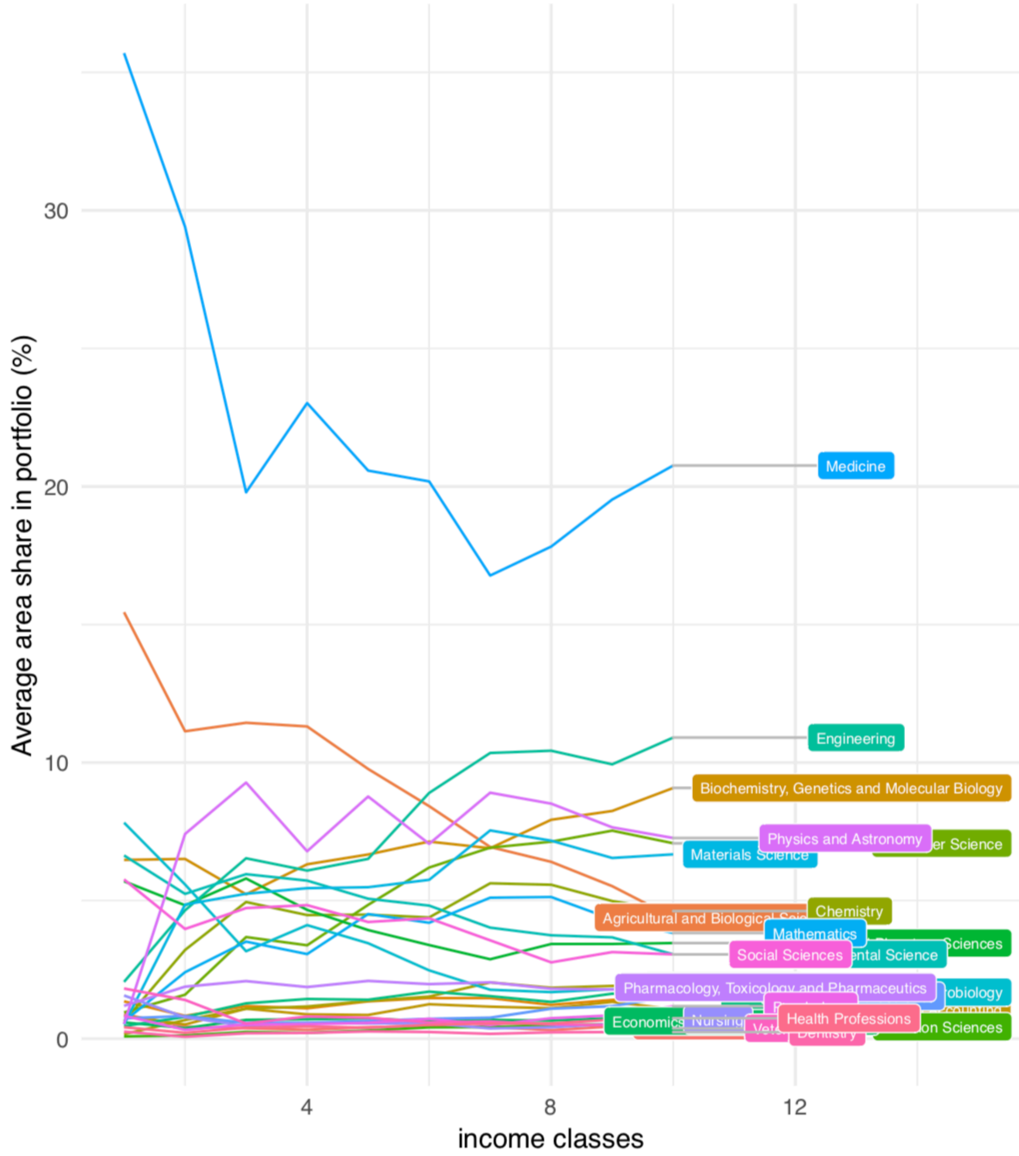


Figure 15. Share of areas in portfolio per entropy class.

Stages in development
Hypothesis 3 focuses on the composition of the research portfolio and the expectation that this may vary over time along the different stages of the economic development trajectory that countries follow. As illustrated in figure 15, showing the average share of research areas in a countries' portfolio, there are some trends and differences to be noticed.

The first thing that may be noticed is that figure 15 can also be used to interpret the earlier discussed results of the clear trend of a decrease of portfolios in terms of the Hoover specialization coefficient as economic growth increases. Here it can be seen that the higher specialization in lower income countries can be attributed to the fact that the 7 largest areas on average take up 83.54% of the portfolio in the lowest income class. Especially Medicine takes up a large share of the portfolio in the lowest income class. Also Agricultural and Biological Sciences, Immunology and Microbiology, Environmental Science, Biochemistry, Genetics and Molecular Biology, Social Sciences and Earth and Planetary Sciences all take up at least 5%. On the contrary in the highest class the distribution of the portfolio is spread over more different fields. The 7 largest fields in this class on average take up 66% of the portfolio, leaving room for a more diversified portfolio of different areas.

Focusing on trends of specific areas the following distinction can be made. Some areas take up a lower share in the middle income countries: Medicine takes up the largest share in every class, but it is remarkable that the lowest two classes have an especially high share, then 3, 4, 5 and 10 are all around or above 20% but there is a drop for the higher middle classes 7, 8 and 9 which have a share lower than 20%.

The fact that some areas have a steady share over all fields may be an indication that increases or decreases in other areas matter. For instance *Pharmacology, Toxicology and Pharmaceutics* and *Dentistry* both have a remarkably similar share in all fields, around 2% and 0.2% respectively.

On the other hand some areas drop in share over higher income classes countries; *Agricultural and Biological sciences* drops from a share of 15,4% in the countries at the lowest entropy class, to only about 4.3% in the countries in the highest income class. *Social sciences* drops in share as well from about 5.7% to only 3% in higher income classes. *Environmental science*, and *Earth and Planetary science* drop in share as well (6.6% to 3% and 5.7% to 3.4%).

On the contrary other areas show a remarkable increase as countries have a higher GDP: both *Materials science* and *Computer science* increase from below 1% in the lowest class to about 7% in the highest class. The fact that these fields have grown a lot over the last two decades in absolute global scientific output, combined with this insight in which income class countries they have the highest share in the research portfolio, says something about both the complexity of the discipline and the types of countries that can develop knowledge in these areas. A similar increase trend can be seen for *Biochemistry, Genetics and Molecular Biology, Energy, Neuroscience*.

Remarkable is also that some areas seem to have higher shares in middle income class countries such as *Chemistry, Physics and Astronomy* and *Engineering.* This inverted u-shape trend may be in line with successfully developing countries in the Asiatic Region and Eastern Europe developing higher shares in areas like engineering, computer science and physics, and lower shares in medicine.

Precise interpretation of all the data underlying these trends takes a lot of effort as each income class contains a lot of countries and portfolios, and each scientific area contains a lot of different fields. However, clear patterns seem to arise from the data when economic development is linked to the growth and relative share of different scientific areas.

Conclusions on hypothesis 3
Exploring trends in the spatial variation and growth of specific scientific areas and fields has been illustrated to be a fruitful analysis; the expected connection between different stages in economic development and composition of the research portfolio can be confirmed. Indeed there are differences in portfolio composition over different stages of development and over different world regions, and fields or areas may play a different role in different stages of economic development. While it is noticeable that in any stage a few scientific areas together take up the largest share of the portfolio. But the precise shares they hold in countries' portfolios still shows spatial variations.

From the analyses on growth factors of different scientific research areas and their distribution over different parts in the world, we can conclude that there is a difference in portfolio composition in terms of growth potential of the content. Also it seems that different countries have a different capability or willingness to produce certain types of knowledge in different scientific disciplines and with different growth rates.

This may be partly due to country specific context but also with the idea that internal dynamics influence how research knowledge is developed. The results also show that there are differences in fields in terms of specialization or diversification. Some fields seem to be easier to enter without related knowledge than others and can therefore play a different role in a countries knowledge development.

When these results are combined and connected to the earlier found correlations of portfolio characteristics and economic growth they provide a more clear picture of how knowledge portfolios develop over time and what portfolio content determines the found trends and correlations.

Lastly, results should also be interpreted with caution, as biasses in the used data, such as overrepresentation of some scientific disciplines or some countries may also explain or reinforce some of the found trends.

Type of economy and governance quality

While specific country context related to knowledge development may entail many factors and science systems may be complex, it is interesting to see whether we can find out a bit more about the origin of variations in knowledge development between countries, and within country variations over time.

Hypothesis 4 focuses on potential country level constraining or enabling factors for knowledge development; the possible influence of governance quality and type of economy or institutions that define a country. The following analyses aim to provide more insight.

Varieties of capitalism and different types of economies

As described by Hall (2001) countries can be categorized in two different types of capitalistic economies; either liberal market economies or coordinated market economies. These countries differ amongst others in the way firms interact with each other and other actors. An extended categorization based on the idea that there are different types countries based on their type of economies, institutions and business systems is proposed by Witt et al. (2017). When using this taxonomy, nine clusters of countries can be discerned, including the original two types of capitalism. The different countries in these clusters are presented in figure 16 and table 17.

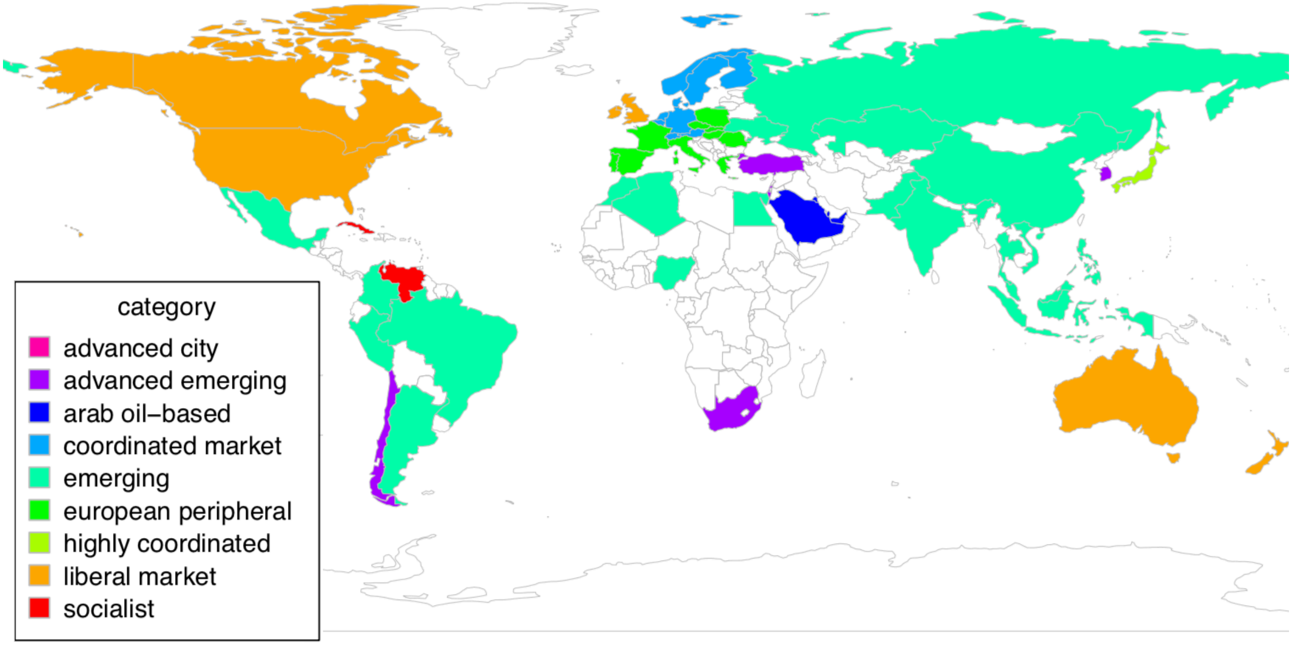


Figure 16. - Nine different types of economies (Witt et al., 2017).

Socialist economies	Emerging economies	Arab oil-based economies	Advanced city economies	Advanced emerging economies	European peripheral economies	Liberal market economies	Coordinated market economies	Highly coordinated economies
Cuba	Algeria	Kuwait	Hong Kong	Chile	Czech Rep.	Australia	Austria	Japan
Venezuela	Argentina	Qatar	Singapore	Israel	France	Canada	Belgium	
	Bangladesh	S. Arabia		Korea	Greece	Ireland	Denmark	
	Brazil	UAE		S. Africa	Hungary	N. Zealand	Finland	
	China			Taiwan	Italy	UK	Germany	
	Colombia			Turkey	Poland	USA	Netherlands	
	Egypt				Portugal		Norway	
	India				Romania		Sweden	
	Indonesia				Slovakia		Switzerland	
	Kazakhstan				Spain			
	Malaysia							
	Mexico							
	Morocco							
	Nigeria							
	Pakistan							
	Peru							
	Philippines							
	Russia							
	Thailand							
	Ukraine							
	Vietnam							

Table 18. - Nine clusters of types of economies and the corresponding countries in (Witt et al., 2017).

Whether the categorization of countries is legitimate and useful in the context of this study can be determined with an ANOVA test. This shows whether the variation between different clusters in terms of GDP and entropy is bigger than the variation between countries in the different clusters. Table 19 shows the results of the ANOVA test.

Entropy	Degrees of freedom	Sum Sq.	Mean Sq.	F value	Pr(>F)
Group	8	501.57	62.697	45.793	< 2.2e-16 ***
Residuals	1227	1679.91	1.369		
LogGDP	Degrees of freedom	Sum Sq.	Mean Sq.	F value	Pr(>F)
Group	8	73.535	9.1919	103.8	< 2.2e-16 ***
Residuals	1230	108.923	0.0886		

Table 19. - ANOVA test output - response variable Entropy and logGDP. (Sig. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’)

Following the results of these tests the ‘type of economy’ clusters can indeed be seen as legitimate distinct categories of countries, also in terms of knowledge development and economic growth. Also this categorization provides stronger results than when countries are clustered by the earlier used categorization of the world region they are in. This was also tested by comparing other variables than entropy and GDP such as the average growth factor.

Taking the analysis to a level of clusters of countries instead of individual countries has advantages. The fact that this way clustering concerns groups of countries suggests that there may be similarities between countries within clusters in terms of portfolio composition that may result in their specific development trajectory and the fact that they are closer to each other than to countries in other groups.

When plotting the different clusters in an entropy - logGDP plot (see figure 17) we find the following. While some clusters are spread out over the whole figure, it does seem like there is a difference between the clusters, as they are concentrated in different positions of the plot.

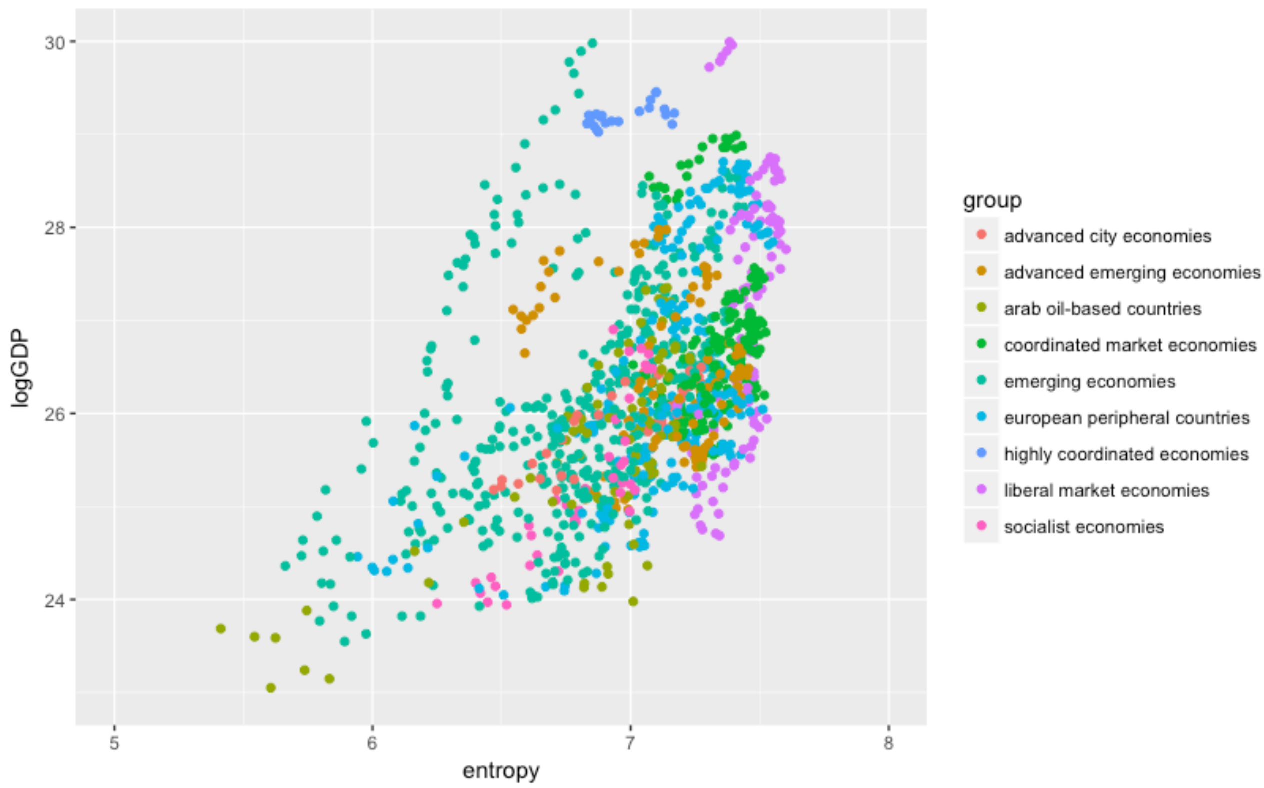


Figure 17. - Position of different economies in terms of economic growth and knowledge development.

When we take a look at the portfolio shares in different growth categories of fields we find the following results (see table 20 and figure 18):

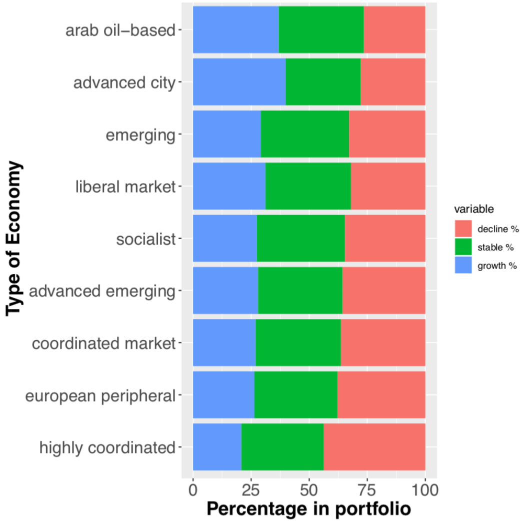


Figure 18. - Share of growth categories.

Cluster	decline	stable	growth	Average growth factor
arab oil-based	21,3	37,7	41,0	1.412146
advanced city	25,9	32,4	41,7	1.518854
Emerging	29,0	38,5	32,5	1.267430
liberal market	30,2	36,9	32,9	1.400705
Socialist	33,0	39,7	27,2	1.189003
advanced emerging	33,3	36,6	30,1	1.257421
coordinated market	34,0	37,2	28,8	1.225416
european peripheral	34,8	36,1	29,2	1.221985
highly coordinated	43,0	35,7	21,3	1.124624

Table. 20. Share in growth categories and growth factor of economies.

This shows that Arab Oil-based countries and Advanced city economies have quite a higher share than the more coordinated economies below in the table. While the amount of ‘stable’ fields is quite similar in different types of economies, the difference lies in the higher shares of fields in ‘decline’ and lower shares in ‘growth’ and vice versa. The difference between coordinated market economies and liberal market economies is only a few percentages. But the expected difference; that liberal markets are better at enabling emerging fields to grow, seems to be confirmed.

When the economies are compared by average growth factor it can also be seen that Advanced city economies, Arab oil-based, Advanced city and Liberal market economies have on average the highest growth factor fields in their portfolio. On the other hand Socialist, Highly coordinated and Coordinated market economies have lower average growth factor fields in their portfolio. This may mean that some types of economy and corresponding institutions are less appropriate for fast growing scientific disciplines to grow than others.

In order to test whether there is a significant difference between country clusters, a between estimator model is used as presented in table 21. Note that the intercept can be used to compare to Liberal market economies - the baseline. Type of economy is the independent variable, and as it is categorical the estimates represent whether the growth factor is significantly higher or lower than the base line.

This shows that there are no significant differences between countries with Arab Oil-based economies and Advanced city economies compared to Liberal market economies. This could be expected as the average values are quite similar. However, the test also shows that the other types of economies, more coordinated ones and the emerging and socialist economies, have fields with a significantly lower growth factor.

Model	IV	Effect (+ significance)	DV	R^2	N
24	Intercept	**1.400705 (< 2.2e-16) *****	Growth factor	0.4943	1239
	advanced city	0.118148 (0.1062709)			
	advanced emerging	**-0.143284 (0.0096917) ****			
	arab oil-based	0.011441 (0.8411486)			
	coordinated market	**-0.175289 (0.0004182) *****			
	emerging	**-0.133275 (0.0020393) ****			
	european peripheral	**-0.178720 (0.0002580) *****			
	highly coordinated	**-0.276081 (0.0054494) ****			
	socialist	**-0.211703 (0.0048596) ****			

Table 21. - PLM regression analysis results summary - DV growth factor.

Portfolio composition of different types of economies

Before diving deeper in possible underlying explanations for this difference, the specific content of portfolios of these different types of economies is explored. Taking a look at portfolio composition in terms of shares of different research areas (see figure 19, next page) it is found that some areas are more present in specific clusters than others. A few noticeable differences are summarized in table 22.

What stands out is that some types of economies on average have lower shares of medicine and higher shares of other areas in their portfolio. Advanced city, Emerging, Advanced emerging, and Highly coordinated economies seem to have higher shares in areas as *engineering, materials science, computer science* and *chemistry.* Which may explain the lower share in *medicine,* as their focus is spread over more fields. Other remarkable points are for instance the high share of *agricultural and biological sciences* in Socialist economies.

However, it is hard to discover further clear trends, and the figure does not reflect absolute amounts of publications produced by different types of countries. Taking another approach, plotting the share of the total growth of scientific areas accounted for by the different clusters of types of economies for the whole period of 1996-2016, figure 20 is produced. This illustrates the role the different types of economies play in the global science production and also how publication efforts of different types of economies are skewed towards certain scientific areas.

Acknowledging that the graphs for Arab oil-based and Socialist economies are hard to read, and only a few spikes of the Advanced cities and Advanced emerging economies can be discerned, the data behind these graphs is still interesting to see to what extent these economies focus on certain fields. On the other hand for the other five types of economies clear trends and differences can be seen.

Most remarkable is the fact that in almost all areas liberal market economies together contributed to the highest share in growth. What is also interesting is that there seems to be a clear thematic preference for certain medical related science disciplines such as *Psychology, Nursing, Health professions, Medicine, Neuroscience etc.* And some typical areas as *Social sciences* and *Arts and Humanities.* And economic disciplines as *Business, Management and Accounting, Economics, Econometrics and Finance.*

None of the other economies shows a clear tendency to the research areas on the right. On the contrary, emerging economies contributed to a large share of the growth of the areas on the left, with the exception of a high spike in the share of multidisciplinary research. The areas on the left also seem to be a cluster of thematically similar scientific areas such as *Materials science,*

Advanced city	Advanced emerging	Arab oil based	Coordinated market	Emerging	European Peripheral	Highly coordinated	Liberal market	Socialist
Higher engineering and computer science share than medicine	High in medicine, engineering, biochemistry, materials science, physics and chemistry	Highest share in energy, high in engineering and computer science	High in medicine, biochemistry, engineering, physics and materials science.	Higher engineering share than medicine, relatively high in physics and material science	High in medicine, biochemistry, engineering and physics.	High in engineering and biochemistry. Relatively high physics and material science.	High in medicine, highest in social science	Agriculture 2nd share after medicine, biochemistry engineering and physics big too.

Table 22. - Overview of portfolio composition features of different economies.

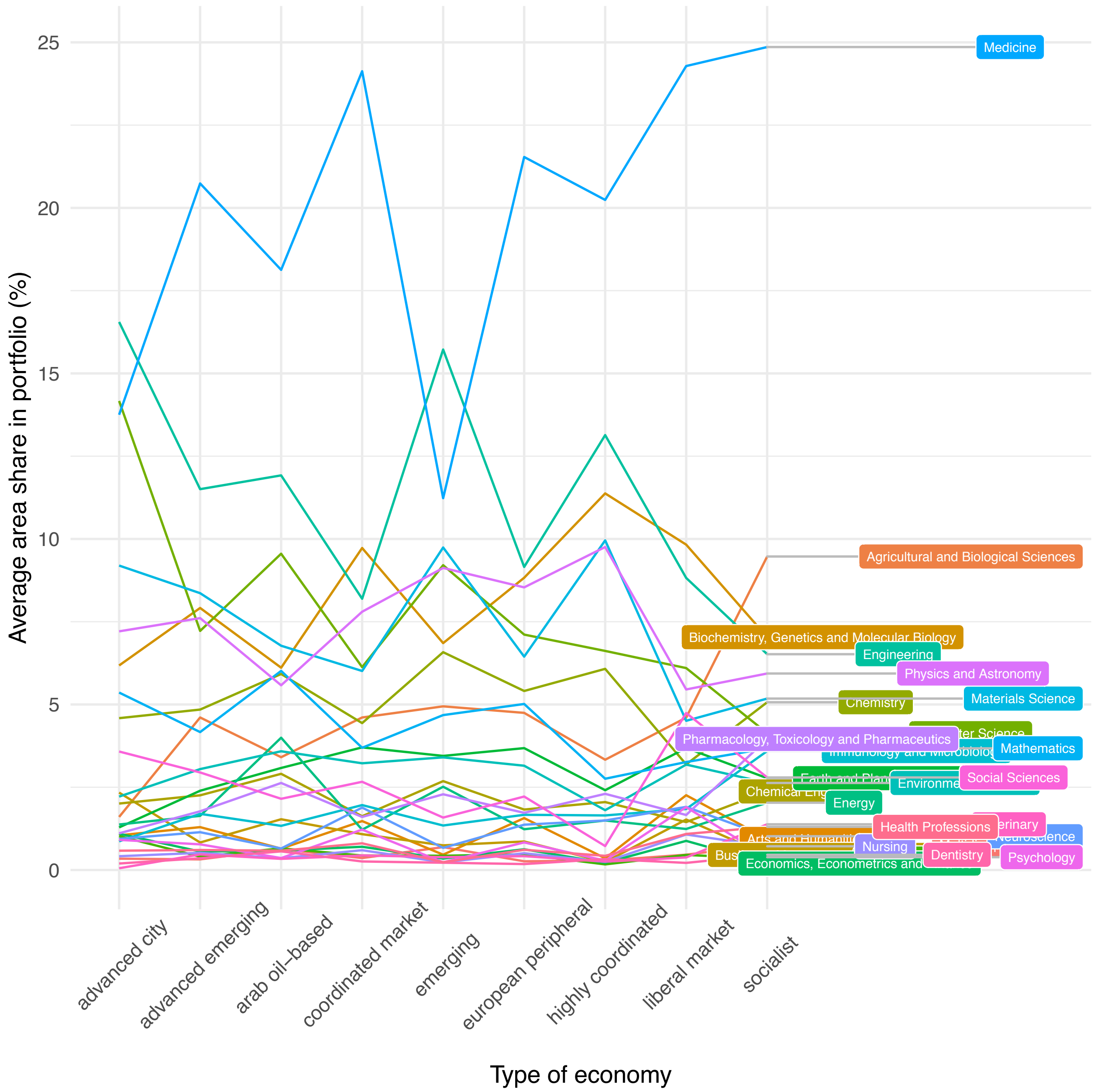


Figure 19. - Average portfolio composition in terms of scientific areas over different types of economies.

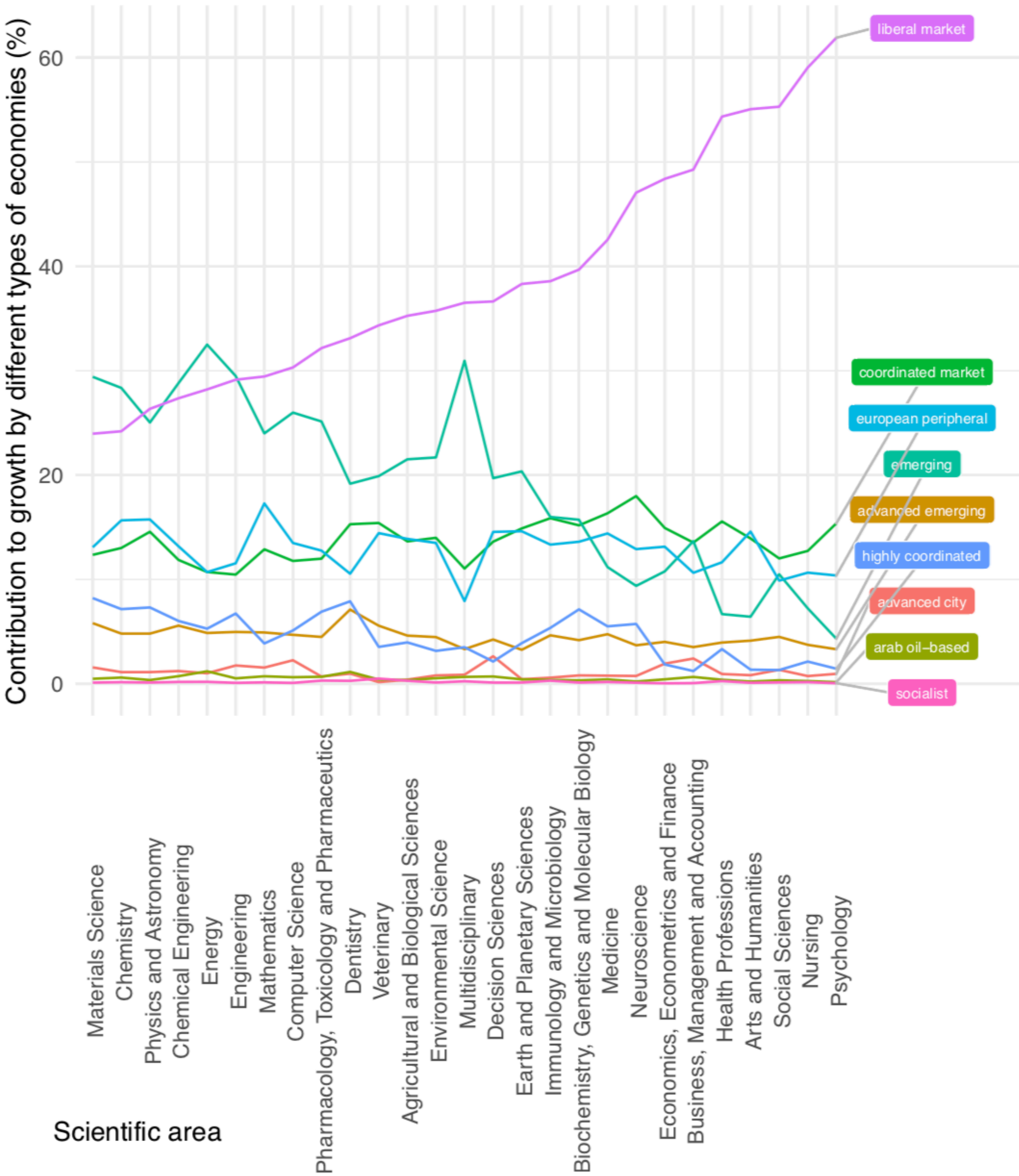


Figure 20. - Share of growth taken up by different types of economies - whole period 1996 to 2016.

Chemistry, Physics and Astronomy, Chemical engineering, Energy, Engineering and *Mathematics.* Those all have a strong foundation in natural sciences and are partly more traditional sciences, but also sciences that have emerged a lot in the past two decades such as computer science, energy and engineering.

This major distinction can be interpreted as a clear shift in global knowledge production, from the traditional dominance of English language, Liberal market economies in certain scientific disciplines, to the upcoming Emerging economies in the world which clearly contribute more to very different emerging fields.

Other remarkable differences are for instance that Highly coordinated economies (Japan) contribute most to certain areas that also seem related: *Dentistry, Biochemistry, Genetics and Molecular Biology, Medicine* and *Neuroscience.*

It can also be noticed that the graphs of Coordinated market economies and European Peripheral economies are quite similar. European countries thus seem to contribute to a similar extent to growth in different scientific areas. A few exceptions lie for instance in the higher share of European Peripheral economies in *Mathematics* and the higher share of Coordinated market economies in *Neuroscience*.

Governance and institutions as enabling or restricting factors

As described before, in hypothesis 4, it can be expected that country context matters and that institutional differences and differences in governance quality matter for knowledge development. The results above indicate that content of portfolios and growth rates seem to differ for different types of economies characterized by different institutions. In order to test whether different aspects of governance quality play a role, a number of regression tests are performed.

First, as presented in table 23, the correlation between different governance quality indicators and knowledge complexity, as reflected by entropy, is estimated. This is done with a panel linear model regression with fixed effect for country and year. Due to data availability this was tested for 196 countries and 11 years of panel data. The R^2 value also indicates that the explanatory value of the model is comparable to the earlier models with entropy as a dependent variable. Full output of the regression model can be found in appendix E.

This shows that the knowledge complexity within countries over time is positively correlated with *Control of corruption, Stability and absence of violence and terrorism, Regulatory quality* and V*oice and accountability.* This could be interpreted as follows. If these characteristics are improved in a country over time this has a positive effect on knowledge development, in terms of higher amounts and variety of knowledge produced (as reflected by entropy of the portfolio). Degrading of these characteristics could have a negative impact. The other way around could also be true, knowledge development may have positive influence on the economic development and also on the governance quality in a country.

IV	Effect (+ significance)	DV	R^2	N
Government Effectiveness	-0.1224977 (0.0114969) *	Entropy	0.40851	3409
Control of corruption	0.1039579 (0.0251311) *			
Stability and absence of violence and terrorism	0.0878272 (0.0004454) ***			
Rule of law	-0.1549039 (0.0034258) **			
Regulatory quality	0.1240954 (0.0036956) **			
Voice and accountability	0.1716377 (4.188e-05) ***			

Table 23. - PLM regression analysis WGI indicators vs. entropy

On the other hand *Government effectiveness* and *Rule of law* are negatively correlated with knowledge complexity. This is a remarkable effect. It might be expected that improving governance quality would always have a positive effect on a country and its knowledge development. An explanation might be provided in line with the distinction between liberal and coordinated markets. If a countries government is effective and rule of law is high, this could result in higher coordination and less flexibility, a less liberal market. Through this mechanism those aspects of governance quality could have a negative impact on knowledge development, as they could be restrictive instead of enabling or stimulating.

To explore whether differences between types of economies in governance quality can provide more insight in their knowledge development regression analyses were performed with a between estimator, as presented in table 24. For each world governance indicator a model was estimated in which the types of economies were included as a categorical independent variable and the governance indicator as the dependent variable. The baseline is again set at Liberal market economies. Note that each row in the table represents the results of one model.

In order to provide more insight in some other aspects of the types of economies and the differences between them, models with entropy, GDP and specialization coefficient as dependent variable were also estimated. Note that the estimates can be interpreted as a significant increase or decrease as compared to the coefficient of the intercept - the value of Liberal market

	Intercept	advanced city	advanced emerging	arab oil-based	coordinated market	emerging	european peripheral	highly coordinated	socialist	R^2	N	n
DV - World governance indicators												
Government Effectiveness	1.70 ***	0.20	**-0.87** ***	**-1.30** ***	0.18	**-1.90** ***	**-0.92** ***	-0.28	**-2.40**	0.84	1062	59
Control of corruption	1.85 ***	0.09	**-1.23** ***	**-1.36** ***	0.17	**-2.39** ***	**-1.30** ***	-0.48	**-2.29** ***	0.87	1062	59
Stability and absence of violence and terrorism	0.93 ***	0.16	**-1.24** ***	-0.46	0.24	**-1.69** ***	-0.32	0.11	**-1.28** **	0.70	1062	59
Rule of law	1.73 ***	-0.20	**-1.06** ***	**-1.30** ***	0.08	**-2.20** ***	**-0.95** ***	-0.39	**-2.89** ***	0.89	1062	59
Regulatory quality	1.69 ***	0.24	**-0.86** ***	**-1.36** ***	-0.08	**-1.92** ***	**-0.77** ***	-0.63	**-3.02** ***	0.89	1062	59
Voice and accountability	1.39 ***	**-1.21** **	**-0.82** **	**-2.37** ***	0.11	**-1.84** ***	-0.41	-0.37	**-2.55** ***	0.83	1062	59
DV - knowledge & economy												
Entropy	7.45 ***	**-0.47** *	-0.30	**-0.60** ***	-0.11	**-0.66** ***	**-0.33** *	-0.47	**-0.63** **	0.52	1239	59
LogGDP	27.5 ***	-1.58	-1.09	**-1.95** **	-0.70	**-1.35** *	-1.14	1.70	**-2.28** *	0.27	1236	59
Specialized portfolio	0.17 ***	0.09	0.07	**0.16** **	-0.02	**0.16** ***	0.02	0.01	**0.17** **	0.56	1239	59

Table 24. - regression analysis results summary - DV: WGI & other variables.

economies. The results show that Advanced city, Coordinated Market and Highly coordinated economies do not significantly differ from Liberal market economies in terms of governance quality.

For the other economies; Advanced emerging, Arab oil-based, Emerging and Socialist economies, a general trend can be seen in the results. All 6 aspects of governance quality are significantly lower for these countries, and also entropy and GDP are significantly lower, and the specialization rate of the portfolio is higher than in Liberal market economies, with the exception of Advanced emerging economies.

While the results are not particularly useful to explain the variance between liberal market and coordinated market economies. They could provide an idea of the enabling role governance quality may have. The types of economies with a lower quality also contribute lower shares in absolute scientific output than liberal economies, and their knowledge complexity is lower. This might indicate that a certain level of governance quality is necessary for knowledge development to flourish.

When results are compared it is remarkable that while governance quality of Arab oil based economies, as well as their GDP, is significantly lower, their portfolios still have on average fields with a high growth rate. Similarly while Emerging economies have lower GDP and governance quality, they still exceed the contribution to growth of Liberal market economies in some scientific disciplines.

Conclusions on hypothesis 4
The results show that clustering countries in types of economies can provide insight in the difference institutional context makes for composition of a countries portfolio. It provides a means to explain differences between countries rather than only within countries.

In line with hypothesis 4 it can be confirmed that there are differences in portfolio composition in terms of publications in different areas and in terms of growth factors of different fields. This may reflect that different types of economies are better at, or choose to, develop different kinds and quantities of knowledge. At least it is clear that there are major differences in terms of the contribution to the share of growth of different scientific areas over the last two decades, that different types of economies have made.

Lastly, the regression analyses show that governance quality is also related to knowledge development. Different aspects of governance quality seem to restrict or enable development of more complex knowledge within countries over time. Looking at between country, or between cluster, variance in governance quality seems to suggest that there is a distinction between certain types of economies. Some with higher quality governance produce similar knowledge complexity, while those with lower quality produce lower variety and amounts of publications.

5. Discussion & Conclusion

In this final chapter first a discussion is provided of the theoretical implications of the found results, of possible the limitations and practical implications of this research, as well as suggestions for future research. This is concluded with final conclusions in section 5.2

5.1 Discussion

Theoretical implications

Different strands of literature were used to create a theoretical framework and formulate specific expectations on the development of national research portfolios. Overall, it can be concluded that many of the theoretical mechanisms from other studies were found back in the exploration of publication data and connections with other types of data.

The results of this study provide an addition to the body of scientometric literature which describes the development of science systems. First of all it provides more insight in the networked character and content of science portfolios than many studies that look at country development. Furthermore it provides more insight in the way science systems develop over time as studied by Horlings & van den Besselaar (2011). The results of this study reinforce the trend that they found which compared publication amounts with the variety of areas in which a country publishes. Furthermore their finding that countries can be clustered based on portfolio content can be confirmed. The results show that indeed different kinds of countries, in terms of institutional context, stage in economic development and world regions, have varying portfolios in terms of specific scientific disciplines.

Important inspiration was drawn from studies which measure knowledge through export data or patents, in terms of theoretical mechanisms and measures used. While the methodology is in some aspects similar to those studies, this research aimed to explore and explain different phenomena, and whether similar trends and patterns would be seen as expected was still to be confirmed. However, the results make that the application of network measures and knowledge space theoretical framework to publication data in order to explore and test expectations has been successful. The results confirm that specialization, related diversification and unrelated diversification, based on the theoretical concepts of branching, adjacent possibilities and the cumulative nature of knowledge development, also apply to knowledge development in national science systems.

Furthermore the link between knowledge portfolios and economic development has also been illustrated, both in explaining within country variation over time as in comparing countries or clusters of countries. The results can also provide an addition or specification to the finding of Van Elk et al. (2015) that there is no uniform relationship between scientific knowledge development and economic productivity. While there may not be a uniform relationship in the sense that this is the same for every country, there seem to be a number of portfolio characteristics that have a significant effect on economic productivity, within and across countries. The connection between countries in different stages in economic development, and different corresponding levels of knowledge complexity, that Pinheiro et al. (2018) illustrated, is also confirmed in this study.

Looking at the illustrated specialization and diversification patterns, the fact that some fields of knowledge seem to play a higher role in unrelated diversification than others suggests different that scientific fields develop in different ways. This is reinforced by the fact that specific fields have differing growth rates, and are concentrated in different types of countries. This confirms the findings of Heimeriks & Balland (2015) and ideas of Bonaccorsi (2008) that fields of knowledge have different characteristics and therefore require different smart specialization strategies. The fact that some groups of countries play a major role in the growth of disciplines over the last two decades also implies that country level decisions to invest in a certain research area can influence the global science system.

Another addition to literature is the fact that this study contributes to empirical evidence confirming the Hall (2001) Varieties of Capitalism hypothesis on Liberal market economies versus Coordinated market economies. This study shows that different types of economies have differing portfolio content and specialize in specific research areas. Liberal market economies also develop on average fields with a higher growth rate. It can be interpreted that liberal markets indeed enable growth of certain fields while to much coordination may restrict emerging fields. This study illustrates that the distinction of countries in liberal and coordinated, and also in other types of economies, can be meaningful for scientific knowledge development too. However, a similar clear distinction between unrelated diversification in liberal market economies and related diversification in coordinated market economies such as in the study of Witt & Jackson (2016) was not found.

Lastly, the exploration of the role that different types of economies and governance quality seem to play in knowledge development provided two interesting results. The distinction between positive correlation of four world governance indicators on knowledge complexity reinforce the idea that governance quality can enable certain types of knowledge development. This could also be interpreted from the analysis comparing the governance quality and the scientific output of different types of economies in different fields. On the other hand two world governance indicators had a negative correlation with knowledge complexity could indicate that indeed too much coordination could restrict knowledge development.

Limitations

This study includes several limitations that are important to take into account in the interpretation of the results and its implications. First there are several limitations or consequences of the type of data that was used for this research. Next there are also limitations to some of the methodological choices that have been made.

As discussed in the methods publication data from SCImago JR was used as the main data for this study, based originally on Scopus citation data. While it was argued that the Scopus database provides the highest coverage and quality in citation data of the considered databases, it does induce some limitations to the interpretation of the results. For instance, as mentioned by Jacso (2009) Scopus has a 34% omission rate of country identification, and this is not equally spread over disciplines. Note that the omission rate is higher for data before 1996 than for the data used in this study. Another flaw is the lack of transparency in how the database of SJR has been produced exactly from the Scopus database (Mañana-Rodríguez, 2014).

One of the most important limitations is the overrepresentation of English language journals in the database, and the overrepresentation of certain fields. A solution to this may be to use additional information on the specific over-representations where available. For instance the comparison of coverage of different disciplines of Harzing & Alakangas (2016) could be taken into account. Other examples of altered interpretation due to these biasses could be as follows. For instance the high contribution to growth of Liberal market economies in all fields of science could be attributed to the fact that those countries are native English and therefore their publications would have higher inclusion rates in the database. On the other hand it is also known that some of these countries are the world leaders in terms of absolute scientific output.

However, the fact that Scopus database consists for 15% of non-English language journals does indicate that a lot of research in other languages is included. The question is how much larger this percentage should be to account for all knowledge, but still publications from many countries and languages are included. This indicates that results could still provide interesting insight in the way non-English countries develop knowledge, but interpretation should be done with more caution for possible biasses.

Another limitation induced by the publication database is the existing categorization in areas and fields. While the categorization has been motivated and tested on robustness in studies as While it is a very useful means of studying patterns and variations in development of specific disciplines, it limits the possibility to measure the development of new knowledge, as categories are filling up over time. For instance, for countries that produce knowledge in most fields there are relatively less unrelated new fields left to enter, and diversification rates are automatically lower for those countries. Also it is not possible to inspect publications in more detail than the information that is currently provided in the SJR database on citations, H-index, scientific area and field.

Furthermore the choice for publications as an indicator for knowledge development induces some limitations. This means that other documents as grey literature are not taken into account. As the recent development of alt-metrics suggests different types of indicators may become a valuable additional indicator of knowledge development and could have been used to provide a more complete picture of knowledge developments. As Konkiel (2016) suggests, alt metrics may have limitations now but may very well become more accurate indicators of scientific impact than other bibliometric alternatives as citations. However, the fact that in this study the link between knowledge development and economic impact explored in different ways, this can also provide inspiration for new ways of evaluation scientific impact.

The methodology chosen, using panel data and accordingly chosen statistical methods to analyze the data, limit the generalizability of the findings. The panel data regression model, including fixed effects, implies that found correlations apply within and across nations, but do not explain variance between nations. This should be taken into account in the interpretation of results. An alternative could have been to use between estimator models as was done for some of the governance quality regression tests, however these type of models take time averages and as such lose insight in the dynamics over time and as such fixed effects models were deemed more appropriate for most tests. This does not mean that the used publication data and the results of this study do not enable cross country comparison of developments over time, and of differing content of portfolio. It does however require caution in the interpretation of causality.

Lastly, many methodological choices and choices in data transformation or in the way data was presented can provide biasses or limitations in interpretation or generalizability. However, in order to improve research quality extensive caution and effort was taken to provide as much transparency as possible, to describe every step taken in data preparation and analysis, and to interpret regression results the right way insights and advice of the critical review of Mummolo & Peterson (2018) were taken into account. Also robustness checks were used to control for sensitivity of analyses and to make sure that results were reliable.

Practical implications

It is probably safe to state that there is a lot of information hidden in publication data and with the right tools and methods they can provide a lot of value. In line with the earlier mentioned statement of Klavans & Boyack (2017) it can be concluded that indeed every major stakeholder in the scientific system should be involved with portfolio analysis. This study proves many measures and methods can be used to provide additional insights beyond publication and citation counts.

This study provides methods and measures that enable insight in the differing positions between countries in the knowledge space and how these are related to their economic development. Furthermore the results may be useful for directors and managers in science systems, in order to develop strategies according to their current position and adjacent possibilities. On a nation level it may also provide a new perspective for policy makers on science systems. As both developments over time as differences in content, and consequently based on the current and past knowledge development adjacent possibilities can be identified, this may inform smart specialization or diversification strategies. Insights can be used to inspire or adjust the intended strategies of a country or evaluate effectiveness of measures taken in the past, such as for instance becoming a knowledge economy instead of an economy that almost solely relies on resources exploitation. Also insights in the connection between different types of governance quality, types of economies and the ability to develop certain disciplines or facilitate growth can be useful to adapt strategies to. Thus, looking at knowledge development as related to the type of governance and economy of a country may be useful.

Lastly, a practical implication of the study is that it illustrates how publication data can be used and visualized. When similar and additional measures are used to create tools or dashboards that enable comparison on the level of clusters of countries, knowledge portfolios and specific areas or fields, this may be a valuable addition to the current tools provided by SCImago.

Future research

As this study performed novel methods in a relatively new field of science, many results lead to further questions that may be relevant for future research. A few of those will be elaborated on here.

The limitations in interpretation of analyses provide reason to further explore the variation between countries. Insights in the precise factors that enable or restrict growth of certain scientific disciplines, and understanding in the mechanisms underlying found correlations may help to better understand and influence knowledge development.

Another lead for future research could be the found results on specialization, related and unrelated diversification. Use of other data could provide more insight in the actual flows of knowledge, what exact knowledge is combined in order to enter related or unrelated fields. For instance as suggested by Boschma et al. (2014) citation data could be used to trace whether related fields actually refer back to the expected type of knowledge in the knowledge base, could further specify the role of relatedness.

Lastly, as mentioned earlier scientific knowledge production is a very complex process including different types of actors in all layers of the science system. On top of that there are also interactions with local context and impacts on the economy. The focus on the scientific output on nation level sheds light on one aspect of the knowledge development process. Further research on the trends and correlations found in this research on other levels may provide a more complete understanding scientometric trends. For instance the role of individual universities, or international collaborations could be taken into account. Or further research could explore how developing specific disciplines in science leads to specific economic activities.

5.2 Conclusion

This study aimed to explore knowledge development and how this varies over space and time through the following research question: “*How do national research portfolios vary over time and space and how can this be explained?”.* This was studied by performing an explorative quantitative research design, including deductive aspects to test the 4 proposed hypotheses.

Evaluating the aim of this study, to provide more insight in two dimensions of knowledge development, the spatial variation over countries and the development of specific research disciplines, this has been accomplished. Also the link between knowledge development and possible economic impact as well as enabling or restricting context factors has been illustrated.

Hypothesis 1 could be confirmed, as indeed a distinction can be made in the extent to which country portfolios contain complex knowledge. As conceptualized in the theory section the knowledge space proves to be useful and its characteristics were confirmed in line with hypothesis 2a and 2b. Indeed low income countries have more specialized and less complex knowledge and diversify more into related and unrelated field in the direction of higher economic growth and a corresponding more complex and diversified knowledge portfolio.

Further exploration of the content of portfolios in terms of growth rates and specific scientific disciplines also confirms hypothesis 3. Lastly clustering countries using the distinction between different types of economies, institutions, aspects of governance and varieties of capitalism proves to be useful in further explaining the uneven spatial distribution of knowledge, in line with hypothesis 4.

As the study had an explorative character, not a few but many different characteristics and variables were used in the analyses and there is not one short answer to the main research question on how research portfolios vary over time and space. However the following insights provide an attempt; Research portfolios vary in different stages of economic development in the extent that they are specialized, that they diversify into related or unrelated new research fields, in their size and in their complexity and variety. Furthermore they can be clustered along the type of economies countries are in, the types of knowledge they contain and the extent to which they seem to facilitate growth of specific disciplines.

To conclude, this research has proven to be in a line of enquiry that is both practically relevant as theoretically interesting to explore. Now it is up to future research to further connect the dots.

6. References

Arthur, W. B. (2007). The Nature of Technology: What it is and How it Evolves. New York: The Free Press.

Balland, P. A., & Rigby, D. (2015). The geography and evolution of complex knowledge. Papers in Evolutionary Economic Geography, WP15, 2.

Balland, P. A. (2017). Economic Geography in R: Introduction to the EconGeo package.

Balland, P. A., Boschma, R. A., Crespo, J., & Rigby, D. L. (2017). Smart Specialization Policy in the EU: Relatedness, Knowledge Complexity and Regional Diversification.

Bashir, M. (2013). Bibliometric study of Pakistan's research output and comparison with other selected countries of the world. Asian Journal of Science and Technology, 4(05), 001-007.

Bonaccorsi, A. (2008). Search regimes and the industrial dynamics of science. Minerva, 46(3), 285.

Boschma, R., Balland, P.A. and Kogler, D. (2015) Relatedness and Technological Change in Cities: The rise and fall of technological knowledge in U.S. metropolitan areas from 1981 to 2010, Industrial and Corporate Change 24 (1): 223-250

Boschma, R., & Capone, G. (2016). Relatedness and diversification in the European Union (EU-27) and European Neighbourhood Policy countries. Environment and Planning C: Government and Policy, 34(4), 617-637.

Boschma, R., Heimeriks, G., & Balland, P. A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. Research Policy, 43(1), 107-114.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. Research Policy, 32/1: 69–87.

Bryman, A. (2012). Social research methods. Fourth edition. Pp. 65-66.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. Journal of Business & Economic Statistics, 29(2), 238-249.

Cano-Kollmann, M., Cantwell, J., Hannigan, T. J., Mudambi, R., & Song, J. (2016). Knowledge connectivity: An agenda for innovation research in international business.

Carayannis, E., & Grigoroudis, E. (2016). Quadruple innovation helix and smart specialization: knowledge production and national competitiveness. Форсайт, 10(1 (eng)).

Cicerone, G., McCann, P., & Venhorst, V. (2017). Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space (No. 1714). Utrecht University, Section of Economic Geography.

Cooke, P., & Leydesdorff, L. (2006). Regional development in the knowledge-based economy: The construction of advantage. The journal of technology Transfer, 31(1), 5-15.

David, P. A. (1994). Why are institutions the 'carriers of history'?: Path dependence and the evolution of conventions, organizations and institutions. Structural change and economic dynamics, 5(2), 205-220.

van Eck, N.J. and Waltman, L. (2009) How to normalize cooccurrence data? An analysis of some well-known similarity measures, Journal of the American Society for Information Science and Technology 60 (8): 1635-1651 van Elk, R., Verspagen, B., Ter Weel, B., Van der Wiel, K., & Wouterse, B. (2015). A macroeconomic analysis of the returns to public R&D investments. CPB Netherlands Bureau for Economic Policy Analysis, 313.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008a). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. The FASEB journal, 22(2), 338-342.

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008b). Comparison of SCImago journal rank indicator with journal impact factor. The FASEB journal, 22(8), 2623-2628.

Florida, R. (2005). The world is spiky. Globalization has changed the economic playing field, but hasn't leveled it. Atlantic Monthly, 296(3), 48.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. Regional studies, 41(5), 685-697.

Fu, X., Pietrobelli, C., & Soete, L. (2011). The role of foreign technology and indigenous innovation in the emerging economies: technological change and catching-up. World development, 39(7), 1204-1212.

Gholizadeh, H., Salehi, H., Embi, M. A., Danaee, M., Motahar, S. M., Ale Ebrahim, N., ... & Azuan, N. (2014). Relationship among economic growth, internet usage and publication productivity: comparison among ASEAN and world's best countries.

Gómez-Núñez, A. J., Vargas-Quesada, B., de Moya-Anegón, F., & Glänzel, W. (2011). Improving SCImago Journal & Country Rank (SJR) subject classification through reference analysis. Scientometrics, 89(3), 741.

Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M., Gómez-Crisóstomo, R., & de Moya-Anegón, F. (2007). Import-export of knowledge between scientific subject categories: The iceberg hypothesis. Scientometrics, 71(3), 423-441.

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. Journal of Informetrics, 6(4), 674-688.

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics, 109(3), 1695-1709.

Hall, P. A. (2001). Varieties of capitalism: The institutional foundations of comparative advantage. D. W. Soskice (Ed.). Oxford: Oxford University Press.

Hancock, E. (2016). Network entropy. University of York. Retrieved on 17.07.2018 via http://www.dsi.unive.it/isscn/wp-content/uploads/2016/07/Hancock.pdf

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., & Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. World Development, 93, 75-93.

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics, 106(2), 787-804.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). The atlas of economic complexity: Mapping paths to prosperity. Mit Press.

Heimeriks, G., & Balland, P. A. (2015). How smart is specialisation? An analysis of specialisation patterns in knowledge production. Science and Public Policy, 43(4), 562-574.

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. proceedings of the national academy of sciences, 106(26), 10570-10575.

Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. Science, 317(5837), 482-487.

Hood, W., & Wilson, C. (2001). The literature of bibliometrics, scientometrics, and informetrics. Scientometrics, 52(2), 291-314.

Horlings, E., & Van den Besselaar, P. (2011). Convergence in science: Growth and structure of worldwide scientific output, 1993–2008. In Science and Innovation Policy, 2011 Atlanta Conference on (pp. 1-19). IEEE.

Howells, J. R. (2002). Tacit knowledge, innovation and economic geography. Urban studies, 39(5-6), 871-884.

Ivanova, I., Strand, Ø., Kushnir, D., & Leydesdorff, L. (2017). Economic and technological complexity: A model study of indicators of knowledge-based innovation systems. Technological Forecasting and Social Change, 120, 77-89.

Jackson, G., & Deeg, R. (2008). Comparing capitalisms: Understanding institutional diversity and its implications for international business. Journal of International Business Studies, 39(4), 540-561.

Jacsó, P. (2009). Errors of omission and their implications for computing scientometric measures in evaluating the publishing productivity and impact of countries. Online information review, 33(2), 376-385.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). "The Worldwide Governance Indicators: Methodology and Analytical Issues". World Bank Policy Research Working Paper No. 5430.

Kharabaf, S., & Abdollahi, M. (2012). Science growth in Iran over the past 35 years. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences, 17(3), 275.

Klavans, R., & Boyack, K. W. (2017). Research portfolio analysis and topic prominence. Journal of Informetrics, 11(4), 1158-1174.

Konkiel, S. (2016). Altmetrics: diversifying the understanding of influential scholarship. Palgrave Communications, 2, 16057.

López-Illescas, C., de Moya Anegón, F., & Moed, H. F. (2009). Comparing bibliometric country-by-country rankings derived from the Web of Science and Scopus: the effect of poorly cited journals in oncology. Journal of Information Science, 35(2), 244-256.

Mañana-Rodríguez, J. (2014). A critical review of SCImago journal & country rank. Research evaluation, 24(4), 343-354.

Mêgnigbêto, E. (2012). Scientific Publishing in West Africa: A Comparison of Benin with Senegal and Ghana. In 17th International Conference on Science and Technology Indicators (pp. 589-600).

Moed, H. (2002). Measuring China's research performance using the Science Citation Index. Scientometrics, 53(3), 281-296.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics, 106(1), 213-228.

Mummolo, J., & Peterson, E. (2018). Improving the Interpretation of Fixed Effects Regression Results. Political Science Research and Methods, 1-7.

Nederhof, A., & Van Wijk, E. (1997). Mapping the social and behavioral sciences world-wide: Use of maps in portfolio analysis of national research efforts. Scientometrics, 40(2), 237-276.

Oaxaca, R. L., & Geisler, I. (2003). Fixed effects models with time invariant variables: a theoretical note. Economics Letters, 80(3), 373-377.

Patelli, A., Cimini, G., Pugliese, E., & Gabrielli, A. (2017). The scientific influence of nations on global scientific and technological development. Journal of Informetrics, 11(4), 1229-1237.

Petralia, S., Balland, P. A., & Morrison, A. (2017). Climbing the ladder of technological development. Research Policy, 46(5), 956-969.

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R., & Hidalgo, C. (2018). Shooting Low or High: Do Countries Benefit from Entering Unrelated Activities?. arXiv preprint arXiv:1801.05352.

Powell, K. R., & Peterson, S. R. (2017). Coverage and quality: A comparison of Web of Science and Scopus databases for reporting faculty nursing publication metrics. Nursing outlook, 65(5), 572-578.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rigby, D. L. (2015). Technological relatedness and knowledge space: entry and exit of US cities from patent classes. Regional Studies, 49(11), 1922-1937.

van Rijnsoever, F. J., Van Den Berg, J., Koch, J., & Hekkert, M. P. (2015). Smart innovation policy: How network position and project composition affect the diversity of an emerging technology. Research Policy, 44(5), 1094-1107.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

SCImago, (n.d.). SJR — SCImago Journal & Country Rank [Portal]. Retrieved January 19th 2018, from http://www.scimagojr.com

Shannon, C.E., Weaver, W. (1949) The Mathematical Theory of Communication. Univ of Illinois Press.

Small, H., & Upham, P. (2009). Citation structure of an emerging research area on the verge of application. Scientometrics, 79(2), 365-375.

Soete, L., Schneegans, S., Eröcal, D., Angathevar, B., & Rasiah, R. (2015). UNESCO Science Report. Towards 2030.

Sorenson, O., Rivkin., J., and Fleming, L. (2006) Complexity, networks and knowledge flow. Research Policy 35(7), 994-1017.

Steijn, M.P.A. (2017) Improvement on the association strength: implementing probability measures based on combinations without repetition, Working Paper, Utrecht University.

Vick, T. E., & Nagano, M. S. (2018). Preconditions for Successful Knowledge Creation in the Context of Academic Innovation Projects. Journal of Information & Knowledge Management, 17(01), 1850004.

The World Bank, World Development Indicators (2018a). GDP (current US$). Retrieved on 27.02.2018 from http://data.worldbank.org/indicator/NY.GDP.MKTP.CD

The World Bank, World Governance Indicators (2018b). GDP (current US$). Retrieved on 09.04.2018 from http://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Witt, M. A., & Jackson, G. (2016). Varieties of Capitalism and institutional comparative advantage: A test and reinterpretation. Journal of International Business Studies, 47(7), 778-806.

Witt, M. A., Kabbach de Castro, L. R., Amaeshi, K., Mahroum, S., Bohle, D., & Saez, L. (2017). Mapping the business systems of 61 major economies: a taxonomy and implications for varieties of capitalism and business systems research. Socio-Economic Review, mwx012.

Woolridge, J. M. Econometric Analysis of Cross Section and Panel Data, 2002.

Yin, R. K. (2013). Case study research: Design and methods. Sage publications.

Zacca-González, G., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., & de Moya-Anegón, F. (2014). Bibliometric analysis of regional Latin America's scientific output in public health through SCImago journal & country rank. BMC public health, 14(1), 632.

7. Appendices

A. Overview scientific areas and fields

Below an overview of the categorization of scientific publications in 27 areas and 310 underlying fields, based on co-citation clusters, as performed by Guerrero-Bote & Moya-Anegón (2012).

Areaname	Areacode	Categoryname	Categorycode
Agricultural and Biological Sciences	1100	Agricultural and Biological Sciences (miscellaneous)	1101
Agricultural and Biological Sciences	1100	Agronomy and Crop Science	1102
Agricultural and Biological Sciences	1100	Animal Science and Zoology	1103
Agricultural and Biological Sciences	1100	Aquatic science	1104
Agricultural and Biological Sciences	1100	Ecology, Evolution, Behavior and Systematics	1105
Agricultural and Biological Sciences	1100	Food Science	1106
Agricultural and Biological Sciences	1100	Forestry	1107
Agricultural and Biological Sciences	1100	Horticulture	1108
Agricultural and Biological Sciences	1100	Insect Science	1109
Agricultural and Biological Sciences	1100	Plant Science	1110
Agricultural and Biological Sciences	1100	Soil Science	1111
Arts and Humanities	1200	Archeology (arts and humanities)	1204
Arts and Humanities	1200	Arts and Humanities (miscellaneous)	1201
Arts and Humanities	1200	Classics	1205
Arts and Humanities	1200	Conservation	1206
Arts and Humanities	1200	History	1202
Arts and Humanities	1200	History and Philosophy of Science	1207
Arts and Humanities	1200	Language and Linguistics	1203
Arts and Humanities	1200	Literature and Literary Theory	1208
Arts and Humanities	1200	Museology	1209
Arts and Humanities	1200	Music	1210
Arts and Humanities	1200	Philosophy	1211
Arts and Humanities	1200	Religious Studies	1212
Arts and Humanities	1200	Visual Arts and Performing Arts	1213
Biochemistry, Genetics and Molecular Biology	1300	Aging	1302
Biochemistry, Genetics and Molecular Biology	1300	Biochemistry	1303
Biochemistry, Genetics and Molecular Biology	1300	Biochemistry, Genetics and Molecular Biology (miscellaneous)	1301
Biochemistry, Genetics and Molecular Biology	1300	Biophysics	1304
Biochemistry, Genetics and Molecular Biology	1300	Biotechnology	1305
Biochemistry, Genetics and Molecular Biology	1300	Cancer Research	1306
Biochemistry, Genetics and Molecular Biology	1300	Cell Biology	1307
Biochemistry, Genetics and Molecular Biology	1300	Clinical Biochemistry	1308
Biochemistry, Genetics and Molecular Biology	1300	Developmental Biology	1309
Biochemistry, Genetics and Molecular Biology	1300	Endocrinology	1310
Biochemistry, Genetics and Molecular Biology	1300	Genetics	1311
Biochemistry, Genetics and Molecular Biology	1300	Molecular Biology	1312
Biochemistry, Genetics and Molecular Biology	1300	Molecular Medicine	1313
Biochemistry, Genetics and Molecular Biology	1300	Physiology	1314
Biochemistry, Genetics and Molecular Biology	1300	Structural Biology	1315
Business, Management and Accounting	1400	Accounting	1402
Business, Management and Accounting	1400	Business and International Management	1403
Business, Management and Accounting	1400	Business, Management and Accounting (miscellaneous)	1401
Business, Management and Accounting	1400	Industrial Relations	1410
Business, Management and Accounting	1400	Management Information Systems	1404
Business, Management and Accounting	1400	Management of Technology and Innovation	1405
Business, Management and Accounting	1400	Marketing	1406
Business, Management and Accounting	1400	Organizational Behavior and Human Resource Management	1407
Business, Management and Accounting	1400	Strategy and Management	1408
Business, Management and Accounting	1400	Tourism, Leisure and Hospitality Management	1409
Chemical Engineering	1500	Bioengineering	1502
Chemical Engineering	1500	Catalysis	1503
Chemical Engineering	1500	Chemical Engineering (miscellaneous)	1501
Chemical Engineering	1500	Chemical Health and Safety	1504
Chemical Engineering	1500	Colloid and Surface Chemistry	1505
Chemical Engineering	1500	Filtration and Separation	1506
Chemical Engineering	1500	Fluid Flow and Transfer Processes	1507
Chemical Engineering	1500	Process Chemistry and Technology	1508
Chemistry	1600	Analytical Chemistry	1602
Chemistry	1600	Chemistry (miscellaneous)	1601
Chemistry	1600	Electrochemistry	1603
Chemistry	1600	Inorganic Chemistry	1604
Chemistry	1600	Organic Chemistry	1605
Chemistry	1600	Physical and Theoretical Chemistry	1606
Chemistry	1600	Spectroscopy	1607
Computer Science	1700	Artificial Intelligence	1702
Computer Science	1700	Computational Theory and Mathematics	1703
Computer Science	1700	Computer Graphics and Computer-Aided Design	1704
Computer Science	1700	Computer Networks and Communications	1705
Computer Science	1700	Computer Science Applications	1706
Computer Science	1700	Computer Science (miscellaneous)	1701
Computer Science	1700	Computer Vision and Pattern Recognition	1707
Computer Science	1700	Hardware and Architecture	1708

Computer Science	1700	Human-Computer Interaction	1709
Computer Science	1700	Information Systems	1710
Computer Science	1700	Signal Processing	1711
Computer Science	1700	Software	1712
Decision Sciences	1800	Decision Sciences (miscellaneous)	1801
Decision Sciences	1800	Information Systems and Management	1802
Decision Sciences	1800	Management Science and Operations Research	1803
Decision Sciences	1800	Statistics, Probability and Uncertainty	1804
Dentistry	3500	Dental Assisting	3502
Dentistry	3500	Dental Hygiene	3503
Dentistry	3500	Dentistry (miscellaneous)	3501
Dentistry	3500	Oral Surgery	3504
Dentistry	3500	Orthodontics	3505
Dentistry	3500	Periodontics	3506
Earth and Planetary Sciences	1900	Atmospheric Science	1902
Earth and Planetary Sciences	1900	Computers in Earth Sciences	1903
Earth and Planetary Sciences	1900	Earth and Planetary Sciences (miscellaneous)	1901
Earth and Planetary Sciences	1900	Earth-Surface Processes	1904
Earth and Planetary Sciences	1900	Economic Geology	1905
Earth and Planetary Sciences	1900	Geochemistry and Petrology	1906
Earth and Planetary Sciences	1900	Geology	1907
Earth and Planetary Sciences	1900	Geophysics	1908
Earth and Planetary Sciences	1900	Geotechnical Engineering and Engineering Geology	1909
Earth and Planetary Sciences	1900	Oceanography	1910
Earth and Planetary Sciences	1900	Paleontology	1911
Earth and Planetary Sciences	1900	Space and Planetary Science	1912
Earth and Planetary Sciences	1900	Stratigraphy	1913
Economics, Econometrics and Finance	2000	Economics and Econometrics	2002
Economics, Econometrics and Finance	2000	Economics, Econometrics and Finance (miscellaneous)	2001
Economics, Econometrics and Finance	2000	Finance	2003
Energy	2100	Energy Engineering and Power Technology	2102
Energy	2100	Energy (miscellaneous)	2101
Energy	2100	Fuel Technology	2103
Energy	2100	Nuclear Energy and Engineering	2104
Energy	2100	Renewable Energy, Sustainability and the Environment	2105
Engineering	2200	Aerospace Engineering	2202
Engineering	2200	Architecture	2216
Engineering	2200	Automotive Engineering	2203
Engineering	2200	Biomedical Engineering	2204
Engineering	2200	Building and Construction	2215
Engineering	2200	Civil and Structural Engineering	2205
Engineering	2200	Computational Mechanics	2206
Engineering	2200	Control and Systems Engineering	2207
Engineering	2200	Electrical and Electronic Engineering	2208
Engineering	2200	Engineering (miscellaneous)	2201
Engineering	2200	Industrial and Manufacturing Engineering	2209
Engineering	2200	Mechanical Engineering	2210
Engineering	2200	Mechanics of Materials	2211
Engineering	2200	Media Technology	2214
Engineering	2200	Ocean Engineering	2213
Engineering	2200	Safety, Risk, Reliability and Quality	2212
Environmental Science	2300	Ecological Modeling	2302
Environmental Science	2300	Ecology	2303
Environmental Science	2300	Environmental Chemistry	2304
Environmental Science	2300	Environmental Engineering	2305
Environmental Science	2300	Environmental Science (miscellaneous)	2301
Environmental Science	2300	Global and Planetary Change	2306
Environmental Science	2300	Health, Toxicology and Mutagenesis	2307
Environmental Science	2300	Management, Monitoring, Policy and Law	2308
Environmental Science	2300	Nature and Landscape Conservation	2309
Environmental Science	2300	Pollution	2310
Environmental Science	2300	Waste Management and Disposal	2311
Environmental Science	2300	Water Science and Technology	2312
Health Professions	3600	Chiropractics	3602
Health Professions	3600	Complementary and Manual Therapy	3603
Health Professions	3600	Emergency Medical Services	3604
Health Professions	3600	Health Information Management	3605
Health Professions	3600	Health Professions (miscellaneous)	3601
Health Professions	3600	Medical Assisting and Transcription	3606
Health Professions	3600	Medical Laboratory Technology	3607
Health Professions	3600	Medical Terminology	3608
Health Professions	3600	Occupational Therapy	3609
Health Professions	3600	Optometry	3610
Health Professions	3600	Pharmacy	3611
Health Professions	3600	Physical Therapy, Sports Therapy and Rehabilitation	3612
Health Professions	3600	Podiatry	3613
Health Professions	3600	Radiological and Ultrasound Technology	3614
Health Professions	3600	Respiratory Care	3615
Health Professions	3600	Speech and Hearing	3616
Health Professions	3600	Sports Science	3699

Immunology and Microbiology	2400	Applied Microbiology and Biotechnology	2402
Immunology and Microbiology	2400	Immunology	2403
Immunology and Microbiology	2400	Immunology and Microbiology (miscellaneous)	2401
Immunology and Microbiology	2400	Microbiology	2404
Immunology and Microbiology	2400	Parasitology	2405
Immunology and Microbiology	2400	Virology	2406
Materials Science	2500	Biomaterials	2502
Materials Science	2500	Ceramics and Composites	2503
Materials Science	2500	Electronic, Optical and Magnetic Materials	2504
Materials Science	2500	Materials Chemistry	2505
Materials Science	2500	Materials Science (miscellaneous)	2501
Materials Science	2500	Metals and Alloys	2506
Materials Science	2500	Nanoscience and Nanotechnology	2507
Materials Science	2500	Polymers and Plastics	2508
Materials Science	2500	Surfaces, Coatings and Films	2509
Mathematics	2600	Algebra and Number Theory	2602
Mathematics	2600	Analysis	2603
Mathematics	2600	Applied Mathematics	2604
Mathematics	2600	Computational Mathematics	2605
Mathematics	2600	Control and Optimization	2606
Mathematics	2600	Discrete Mathematics and Combinatorics	2607
Mathematics	2600	Geometry and Topology	2608
Mathematics	2600	Logic	2609
Mathematics	2600	Mathematical Physics	2610
Mathematics	2600	Mathematics (miscellaneous)	2601
Mathematics	2600	Modeling and Simulation	2611
Mathematics	2600	Numerical Analysis	2612
Mathematics	2600	Statistics and Probability	2613
Mathematics	2600	Theoretical Computer Science	2614
Medicine	2700	Anatomy	2702
Medicine	2700	Anesthesiology and Pain Medicine	2703
Medicine	2700	Biochemistry (medical)	2704
Medicine	2700	Cardiology and Cardiovascular Medicine	2705
Medicine	2700	Complementary and Alternative Medicine	2707
Medicine	2700	Critical Care and Intensive Care Medicine	2706
Medicine	2700	Dermatology	2708
Medicine	2700	Drug Guides	2709
Medicine	2700	Embryology	2710
Medicine	2700	Emergency Medicine	2711
Medicine	2700	Endocrinology, Diabetes and Metabolism	2712
Medicine	2700	Epidemiology	2713
Medicine	2700	Family Practice	2714
Medicine	2700	Gastroenterology	2715
Medicine	2700	Genetics (clinical)	2716
Medicine	2700	Geriatrics and Gerontology	2717
Medicine	2700	Health Informatics	2718
Medicine	2700	Health Policy	2719
Medicine	2700	Hematology	2720
Medicine	2700	Hepatology	2721
Medicine	2700	Histology	2722
Medicine	2700	Immunology and Allergy	2723
Medicine	2700	Infectious Diseases	2725
Medicine	2700	Internal Medicine	2724
Medicine	2700	Medicine (miscellaneous)	2701
Medicine	2700	Microbiology (medical)	2726
Medicine	2700	Nephrology	2727
Medicine	2700	Neurology (clinical)	2728
Medicine	2700	Obstetrics and Gynecology	2729
Medicine	2700	Oncology	2730
Medicine	2700	Ophthalmology	2731
Medicine	2700	Orthopedics and Sports Medicine	2732
Medicine	2700	Otorhinolaryngology	2733
Medicine	2700	Pathology and Forensic Medicine	2734
Medicine	2700	Pediatrics, Perinatology and Child Health	2735
Medicine	2700	Pharmacology (medical)	2736
Medicine	2700	Physiology (medical)	2737
Medicine	2700	Psychiatry and Mental Health	2738
Medicine	2700	Public Health, Environmental and Occupational Health	2739
Medicine	2700	Pulmonary and Respiratory Medicine	2740
Medicine	2700	Radiology, Nuclear Medicine and Imaging	2741
Medicine	2700	Rehabilitation	2742
Medicine	2700	Reproductive Medicine	2743
Medicine	2700	Reviews and References (medical)	2744
Medicine	2700	Rheumatology	2745
Medicine	2700	Surgery	2746
Medicine	2700	Transplantation	2747
Medicine	2700	Urology	2748
Multidisciplinary	1000	Multidisciplinary	1000
Neuroscience	2800	Behavioral Neuroscience	2802
Neuroscience	2800	Biological Psychiatry	2803

Neuroscience	2800	Cellular and Molecular Neuroscience	2804
Neuroscience	2800	Cognitive Neuroscience	2805
Neuroscience	2800	Developmental Neuroscience	2806
Neuroscience	2800	Endocrine and Autonomic Systems	2807
Neuroscience	2800	Neurology	2808
Neuroscience	2800	Neuroscience (miscellaneous)	2801
Neuroscience	2800	Sensory Systems	2809
Nursing	2900	Advanced and Specialized Nursing	2902
Nursing	2900	Assessment and Diagnosis	2903
Nursing	2900	Care Planning	2904
Nursing	2900	Community and Home Care	2905
Nursing	2900	Critical Care Nursing	2906
Nursing	2900	Emergency Nursing	2907
Nursing	2900	Fundamentals and Skills	2908
Nursing	2900	Gerontology	2909
Nursing	2900	Issues, Ethics and Legal Aspects	2910
Nursing	2900	Leadership and Management	2911
Nursing	2900	LPN and LVN	2912
Nursing	2900	Maternity and Midwifery	2913
Nursing	2900	Medical and Surgical Nursing	2914
Nursing	2900	Nurse Assisting	2915
Nursing	2900	Nursing (miscellaneous)	2901
Nursing	2900	Nutrition and Dietetics	2916
Nursing	2900	Oncology (nursing)	2917
Nursing	2900	Pediatrics	2919
Nursing	2900	Pharmacology (nursing)	2920
Nursing	2900	Psychiatric Mental Health	2921
Nursing	2900	Research and Theory	2922
Nursing	2900	Review and Exam Preparation	2923
Pharmacology, Toxicology and Pharmaceutics	3000	Drug Discovery	3002
Pharmacology, Toxicology and Pharmaceutics	3000	Pharmaceutical Science	3003
Pharmacology, Toxicology and Pharmaceutics	3000	Pharmacology	3004
Pharmacology, Toxicology and Pharmaceutics	3000	Pharmacology, Toxicology and Pharmaceutics (miscellaneous)	3001
Pharmacology, Toxicology and Pharmaceutics	3000	Toxicology	3005
Physics and Astronomy	3100	Acoustics and Ultrasonics	3102
Physics and Astronomy	3100	Astronomy and Astrophysics	3103
Physics and Astronomy	3100	Atomic and Molecular Physics, and Optics	3107
Physics and Astronomy	3100	Condensed Matter Physics	3104
Physics and Astronomy	3100	Instrumentation	3105
Physics and Astronomy	3100	Nuclear and High Energy Physics	3106
Physics and Astronomy	3100	Physics and Astronomy (miscellaneous)	3101
Physics and Astronomy	3100	Radiation	3108
Physics and Astronomy	3100	Statistical and Nonlinear Physics	3109
Physics and Astronomy	3100	Surfaces and Interfaces	3110
Psychology	3200	Applied Psychology	3202
Psychology	3200	Clinical Psychology	3203
Psychology	3200	Developmental and Educational Psychology	3204
Psychology	3200	Experimental and Cognitive Psychology	3205
Psychology	3200	Neuropsychology and Physiological Psychology	3206
Psychology	3200	Psychology (miscellaneous)	3201
Psychology	3200	Social Psychology	3207
Social Sciences	3300	Anthropology	3314
Social Sciences	3300	Archeology	3302
Social Sciences	3300	Communication	3315
Social Sciences	3300	Cultural Studies	3316
Social Sciences	3300	Demography	3317
Social Sciences	3300	Development	3303
Social Sciences	3300	Education	3304
Social Sciences	3300	E-learning	3399
Social Sciences	3300	Gender Studies	3318
Social Sciences	3300	Geography, Planning and Development	3305
Social Sciences	3300	Health (social science)	3306
Social Sciences	3300	Human Factors and Ergonomics	3307
Social Sciences	3300	Law	3308
Social Sciences	3300	Library and Information Sciences	3309
Social Sciences	3300	Life-span and Life-course Studies	3319
Social Sciences	3300	Political Science and International Relations	3320
Social Sciences	3300	Public Administration	3321
Social Sciences	3300	Safety Research	3311
Social Sciences	3300	Social Sciences (miscellaneous)	3301
Social Sciences	3300	Social Work	3323
Social Sciences	3300	Sociology and Political Science	3312
Social Sciences	3300	Transportation	3313
Social Sciences	3300	Urban Studies	3322
Veterinary	3400	Equine	3402
Veterinary	3400	Food Animals	3403
Veterinary	3400	Small Animals	3404
Veterinary	3400	Veterinary (miscellaneous)	3401

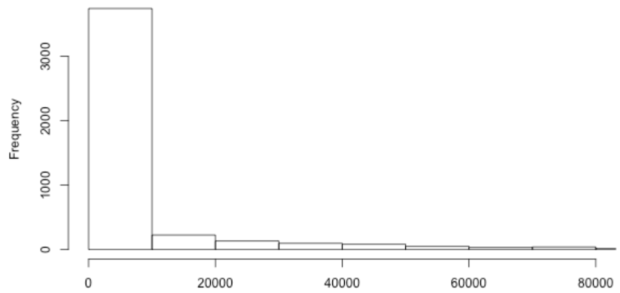
B. Overview countries and world regions

Africa	Asiatic region	Eastern Europe	Latin America	Middle East	Northern America	Pacific Region	Western Europe
Algeria	Afghanistan	Albania	Anguilla	Bahrain	Canada	American Samoa	Andorra
Angola	Bangladesh	Armenia	Antigua and Barbuda	Egypt	Saint Pierre and Miquelon	Australia	Austria
Benin	Bhutan	Azerbaijan	Argentina	Iran	United States	Christmas Island	Belgium
Botswana	Brunei Darussalam	Belarus	Aruba	Iraq	United States Minor Outlying Islands	Cocos (Keeling) Islands	Cyprus
British Indian Ocean Territory	Cambodia	Bosnia and Herzegovina	Bahamas	Israel		Cook Islands	Denmark
Burkina Faso	China	Bouvet Island	Barbados	Jordan		Federated States of Micronesia	Faroe Islands
Burundi	Hong Kong	Bulgaria	Belize	Kuwait		Fiji	Finland
Cameroon	India	Croatia	Bermuda	Lebanon		French Polynesia	France
Cape Verde	Indonesia	Czech Republic	Bolivia	Oman		French Southern Territories	Germany
Central African Republic	Japan	Estonia	Brazil	Palestine		Guam	Gibraltar
Chad	Kazakhstan	Georgia	Cayman Islands	Qatar		Heard Island and McDonald Islands	Greece
Comoros	Kyrgyzstan	Hungary	Chile	Saudi Arabia		Kiribati	Greenland
Congo	Laos	Latvia	Colombia	Syrian Arab Republic		Marshall Islands	Iceland
Côte d’Ivoire	Macao	Lithuania	Costa Rica	Turkey		Nauru	Ireland
Democratic Republic Congo	Malaysia	Macedonia	Cuba	United Arab Emirates		New Caledonia	Italy
Djibouti	Maldives	Moldova	Dominica	Yemen		New Zealand	Liechtenstein
Equatorial Guinea	Mongolia	Montenegro	Dominican Republic			Niue	Luxembourg
Eritrea	Myanmar	Poland	Ecuador			Norfolk Island	Malta
Ethiopia	Nepal	Romania	El Salvador			Palau	Monaco
Gabon	North Korea	Russian Federation	Falkland Islands (Malvinas)			Papua New Guinea	Netherlands
Gambia	Northern Mariana Islands	Serbia	French Guiana			Pitcairn	Norway
Ghana	Pakistan	Slovakia	Grenada			Samoa	Portugal
Guinea	Philippines	Slovenia	Guadeloupe			Solomon Islands	San Marino
Guinea-Bissau	Singapore	Ukraine	Guatemala			Tokelau	Spain
Kenya	South Korea		Guyana			Tonga	Svalbard and Jan Mayen
Lesotho	Sri Lanka		Haïti			Tuvalu	Sweden
Liberia	Taiwan		Honduras			Vanuatu	Switzerland

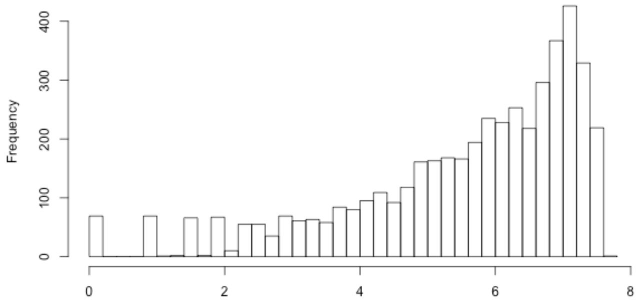
Africa	Asiatic region	Eastern Europe	Latin America	Middle East	Northern America	Pacific Region	Western Europe
Libya	Tajikistan		Jamaica			Wallis and Futuna	United Kingdom
Madagascar	Thailand		Martinique				Vatican City State
Malawi	Timor-Leste		Mexico				
Mali	Turkmenistan		Montserrat				
Mauritania	Uzbekistan		Netherlands Antilles				
Mauritius	Viet Nam		Nicaragua				
Mayotte			Panama				
Morocco			Paraguay				
Mozambique			Peru				
Namibia			Puerto Rico				
Niger			Saint Kitts and Nevis				
Nigeria			Saint Lucia				
Reunion			Saint Vincent and the Grenadines				
Rwanda			South Georgia and the South Sandwich Islands				
Saint Helena			Suriname				
Sao Tome and Principe			Trinidad and Tobago				
Senegal			Turks and Caicos Islands				
Seychelles			Uruguay				
Sierra Leone			Venezuela				
Somalia			Virgin Islands (British)				
South Africa			Virgin Islands (U.S.)				
Sudan							
Swaziland							
Tanzania							
Togo							
Tunisia							
Uganda							
Western Sahara							
Zambia							
Zimbabwe							

C. Histograms - distribution of main variables

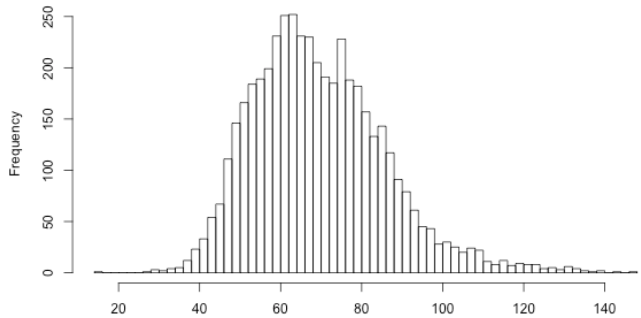
Publication database variables:



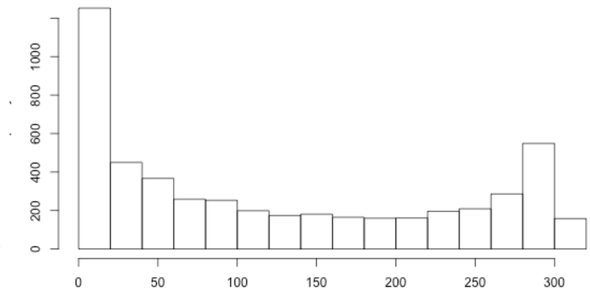
Publication count



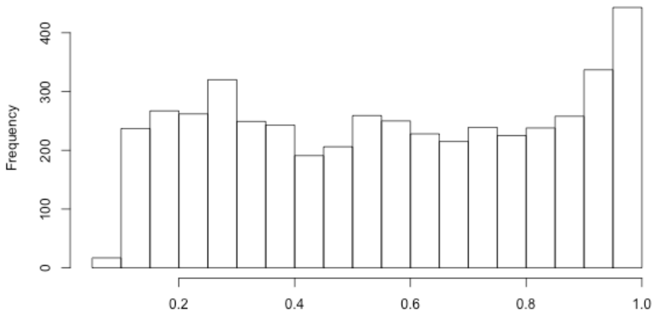
Entropy



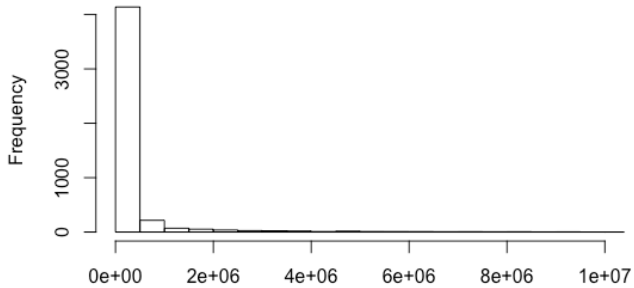
Ubiquity



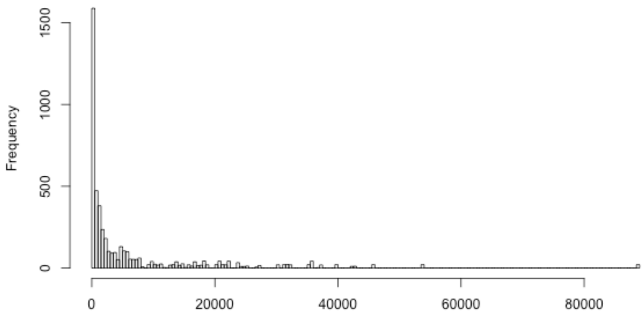
Nr. of fields



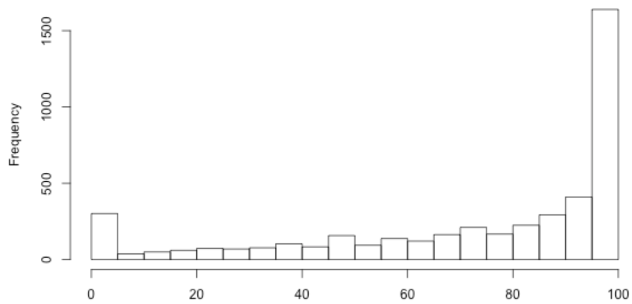
Hoover specialization coefficient



Citations

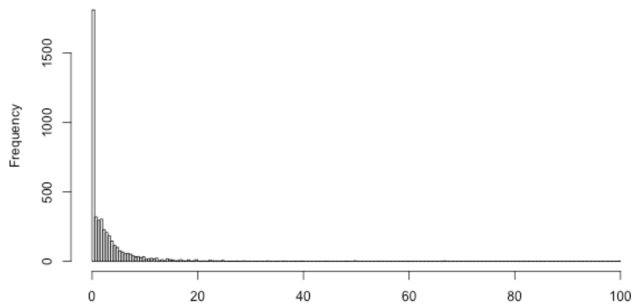


Hirsch index

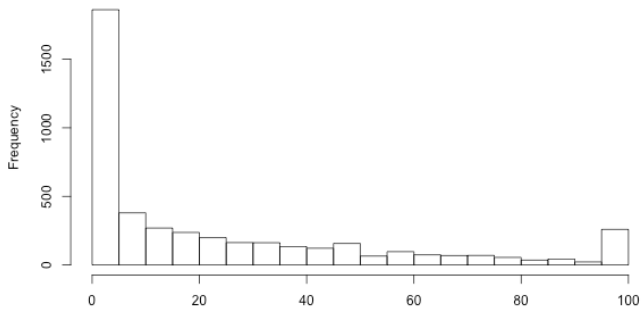


Specialization % (process)

Each of the histograms visualized the distribution of different observations; namely the portfolio of publications of each country for each year.

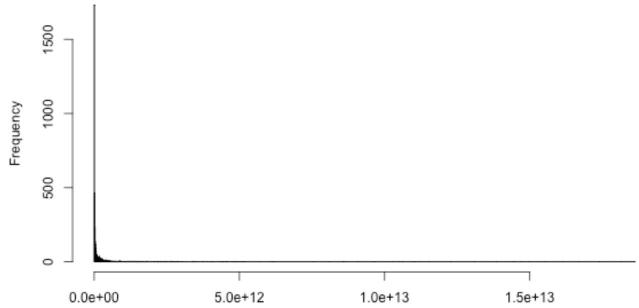


Related diversification % (process)

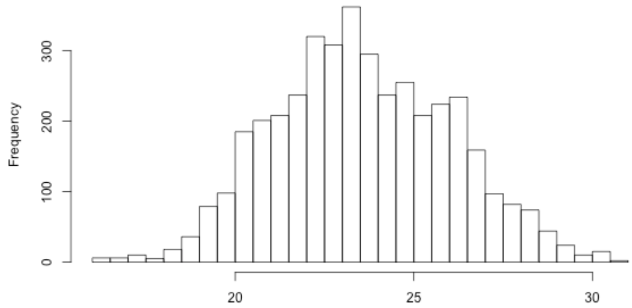


Unrelated diversification % (process)

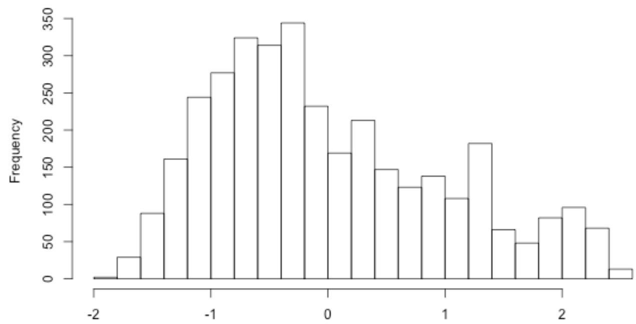
Variables from external sources:



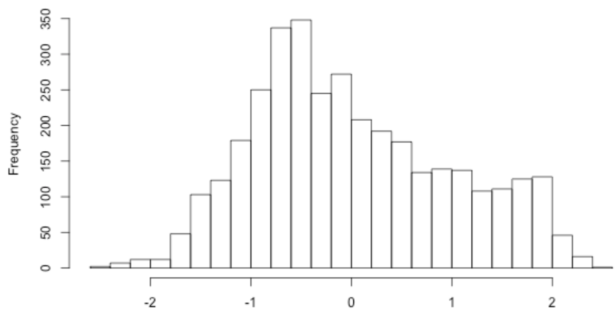
GDP



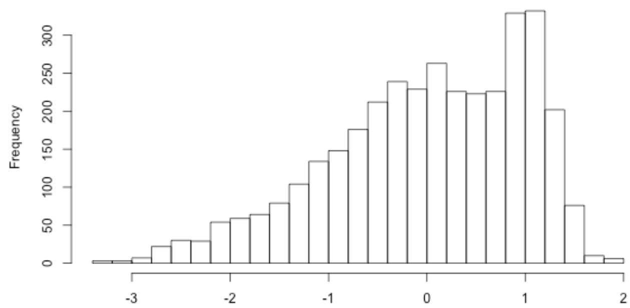
logGDP



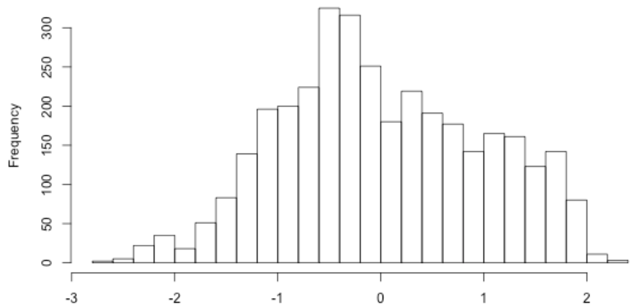
Control of corruption



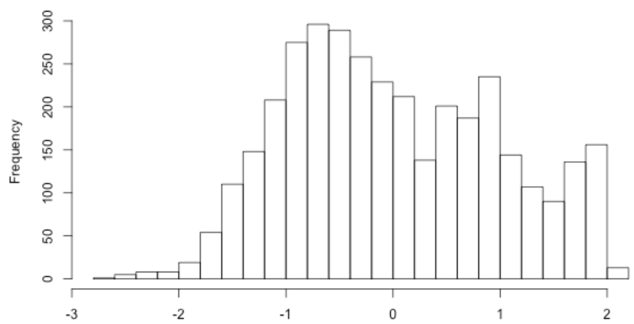
Government effectiveness



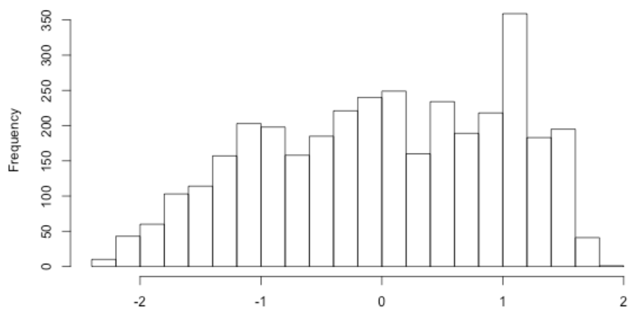
Political stability and absence of violence/terrorism



Regulatory quality



Rule of law



Voice and accountability

Each of the histograms visualizes the distribution of variables on all observations in the data; namely of each country of each year.

D. Assumption tests - motivation of the fixed effect model

This section provides more information on the performed assumption tests which lead to the choice for a fixed effects model such as used in model 1 to 24, and the models presented in table 23.

The Lagrange Multiplier (LM) test was performed in order to test whether an Ordinary Least Squared (OLS) regression was appropriate for the data. The LM test helps to choose between a random effects regression and an OLS regression. The null hypothesis in the LM test is that variances across entities (in this case countries) is zero. This is, no significant difference across units (i.e. no panel effect).

```
plmtest(pooling, type=c("bp"))

        Lagrange Multiplier Test - (Breusch-Pagan) for unbalanced panels

data:  Y ~ X
chisq = 24458, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects
```

-> this proved that there is significant differences between countries and an OLS model should not be used

The F test for individual effects is used to test whether a fixed effects regression would be better suited for the data than an OLS regression.

```
pFtest(fixed, pooling)

        F test for individual effects

data:  Y ~ X + factor(country) + factor(year)
F = 545.19, df1 = 220, df2 = 3776, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Significant results show that it is better to use fixed effects than OLS

The Hausman test is used to compare the fixed effects and the random effects model

```
phtest(fixed,random)

        Hausman Test

data:  Y ~ X + factor(country) + factor(year)
chisq = 3902.4, df = 1, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Significant results show that it is better to use fixed effects than random effects

Lastly an F test is used to test whether fixed effects for the time variable (different years) should be included as well by comparing a model with and a model without a year dummy included.

```
pFtest(fixed.time, fixed)

        F test for individual effects

data:  Y ~ X + factor(year)
F = 426.19, df1 = 20, df2 = 3776, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Significant results show that it is better to control for time too

E. Statistical regression model summaries

A few parts of the first regression output are highlighted in order to facilitate better interpretation of the results.In this section of the appendix first all models such as presented in table 9, 10, 11 & 13 are provided, next the models of table 23 and 24 are provided.

Model 1

Oneway (individual) effect Within Model

Call:
```
plm(formula = logGDP ~ entropy + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))
```

n = the # of countries,
T = the # of years of panel data used in every regression model, and
N = the total # of observations

Unbalanced Panel: n = 201, T = 2-21, N = 3998

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-2.3552067	-0.1428559	0.0047431	0.1482385	1.0408493

Coefficients:

	Estimate	**Std. Error**	**t-value**	**Pr(>ltl)**
entropy	**0.0218543**	0.0089306	2.4471	**0.0144449** *
factor(year)1997	0.0341205	0.0267684	1.2747	0.2025101
factor(year)1998	0.0296372	0.0268080	1.1055	0.2689990
factor(year)1999	0.0454213	0.0268907	1.6891	0.0912808 .
factor(year)2000	0.0882964	0.0267066	3.3062	0.0009547 ***
factor(year)2001	0.1006360	0.0265900	3.7847	0.0001563 ***
factor(year)2002	0.1675117	0.0266101	6.2950	3.425e-10 ***
factor(year)2003	0.3101162	0.0265129	11.6968	< 2.2e-16 ***
factor(year)2004	0.4528186	0.0266425	16.9961	< 2.2e-16 ***
factor(year)2005	0.5768647	0.0267373	21.5753	< 2.2e-16 ***
factor(year)2006	0.7067999	0.0269381	26.2379	< 2.2e-16 ***
factor(year)2007	0.8667501	0.0270133	32.0861	< 2.2e-16 ***
factor(year)2008	1.0158014	0.0270802	37.5108	< 2.2e-16 ***
factor(year)2009	0.9578088	0.0273138	35.0669	< 2.2e-16 ***
factor(year)2010	1.0538278	0.0273772	38.4929	< 2.2e-16 ***
factor(year)2011	1.1688301	0.0275501	42.4256	< 2.2e-16 ***
factor(year)2012	1.2072075	0.0277594	43.4882	< 2.2e-16 ***
factor(year)2013	1.2546573	0.0278098	45.1157	< 2.2e-16 ***
factor(year)2014	1.2894267	0.0281061	45.8772	< 2.2e-16 ***
factor(year)2015	1.2122791	0.0281635	43.0443	< 2.2e-16 ***
factor(year)2016	1.2221189	0.0284356	42.9785	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 1201.1
Residual Sum of Squares: 248.39
R-Squared: **0.79321**
Adj. R-Squared: 0.78111
F-statistic: 689.713 on 21 and 3776 DF, **p-value:** < 2.22e-16

If this number is < 0.05 then the model is ok. This is a test (F) to see whether all the coefficients in the model are different than zero.

Model 2

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ `specialization (% amounts)` + factor(year) +
   factor(country), data = panel.data, model = "within", index = c("country",
   "year"))

Unbalanced Panel: n = 201, T = 1-20, N = 3814

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.0860436 -0.1374027  0.0052867  0.1443573  1.0564468

Coefficients:
                             Estimate  Std. Error t-value  Pr(>|t|)
`specialization (% amounts)`  0.00164799  0.00035916  4.5884 4.618e-06 ***
factor(year)1998            -0.00884301  0.02611011 -0.3387   0.73487
factor(year)1999             0.00643826  0.02619762  0.2458   0.80588
factor(year)2000             0.05208242  0.02599362  2.0037   0.04518 *
factor(year)2001             0.06417224  0.02588675  2.4790   0.01322 *
factor(year)2002             0.13341782  0.02585855  5.1595 2.610e-07 ***
factor(year)2003             0.27523342  0.02574736 10.6898 < 2.2e-16 ***
factor(year)2004             0.41739505  0.02580814 16.1730 < 2.2e-16 ***
factor(year)2005             0.54274602  0.02579319 21.0422 < 2.2e-16 ***
factor(year)2006             0.67354387  0.02582908 26.0770 < 2.2e-16 ***
factor(year)2007             0.83007770  0.02591645 32.0290 < 2.2e-16 ***
factor(year)2008             0.97802553  0.02603661 37.5635 < 2.2e-16 ***
factor(year)2009             0.92380403  0.02598657 35.5493 < 2.2e-16 ***
factor(year)2010             1.01859957  0.02601247 39.1581 < 2.2e-16 ***
factor(year)2011             1.13402200  0.02608764 43.4697 < 2.2e-16 ***
factor(year)2012             1.17126231  0.02626218 44.5988 < 2.2e-16 ***
factor(year)2013             1.21817457  0.02624819 46.4098 < 2.2e-16 ***
factor(year)2014             1.25270951  0.02639838 47.4540 < 2.2e-16 ***
factor(year)2015             1.17393928  0.02651677 44.2716 < 2.2e-16 ***
factor(year)2016             1.18448562  0.02670577 44.3532 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1093.7
Residual Sum of Squares: 224.27
R-Squared:      0.79494
Adj. R-Squared: 0.78238
F-statistic: 696.429 on 20 and 3593 DF, p-value: < 2.22e-16
```

Model 3

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ `rel.div. (% amounts)` + factor(year) +
   factor(country), data = panel.data, model = "within", index = c("country",
   "year"))

Unbalanced Panel: n = 201, T = 1-20, N = 3814

Residuals:
     Min.    1st Qu.     Median    3rd Qu.      Max.
-2.0890714 -0.1368828  0.0024267  0.1423925  1.0269509

Coefficients:
                        Estimate  Std. Error t-value  Pr(>|t|)
`rel.div. (% amounts)` -0.00430309  0.00099679 -4.3169 1.625e-05 ***
factor(year)1998       -0.00604294  0.02610241 -0.2315   0.81693
factor(year)1999        0.01091263  0.02617979  0.4168   0.67682
factor(year)2000        0.05123771  0.02600755  1.9701   0.04890 *
factor(year)2001        0.06557703  0.02589119  2.5328   0.01136 *
factor(year)2002        0.13361195  0.02586695  5.1654 2.530e-07 ***
factor(year)2003        0.27734687  0.02574186 10.7742 < 2.2e-16 ***
factor(year)2004        0.42125490  0.02577775 16.3418 < 2.2e-16 ***
factor(year)2005        0.54813007  0.02574401 21.2916 < 2.2e-16 ***
factor(year)2006        0.67928803  0.02576135 26.3685 < 2.2e-16 ***
factor(year)2007        0.84103875  0.02572570 32.6925 < 2.2e-16 ***
factor(year)2008        0.98742073  0.02585380 38.1925 < 2.2e-16 ***
factor(year)2009        0.93417923  0.02580554 36.2007 < 2.2e-16 ***
factor(year)2010        1.03053671  0.02577195 39.9868 < 2.2e-16 ***
factor(year)2011        1.14729847  0.02582381 44.4279 < 2.2e-16 ***
factor(year)2012        1.18276104  0.02598353 45.5197 < 2.2e-16 ***
factor(year)2013        1.23283789  0.02587915 47.6383 < 2.2e-16 ***
factor(year)2014        1.26892012  0.02595596 48.8874 < 2.2e-16 ***
factor(year)2015        1.19028997  0.02602156 45.7424 < 2.2e-16 ***
factor(year)2016        1.20100577  0.02622207 45.8013 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1093.7
Residual Sum of Squares: 224.42
R-Squared:      0.7948
Adj. R-Squared: 0.78224
F-statistic: 695.843 on 20 and 3593 DF, p-value: < 2.22e-16
```

Model 4

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ `unrel.div (% amounts)` + factor(year) +
   factor(country), data = panel.data, model = "within", index = c("country",
   "year"))

Unbalanced Panel: n = 201, T = 1-20, N = 3814

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.0760661 -0.1379477  0.0035687  0.1451872  1.0550122

Coefficients:
                         Estimate  Std. Error t-value  Pr(>|t|)
`unrel.div (% amounts)` -0.00112089  0.00036498 -3.0711  0.002148 **
factor(year)1998        -0.00684021  0.02614683 -0.2616  0.793639
factor(year)1999         0.00848092  0.02623664  0.3232  0.746527
factor(year)2000         0.05360340  0.02603197  2.0591  0.039553 *
factor(year)2001         0.06519053  0.02592707  2.5144  0.011967 *
factor(year)2002         0.13472514  0.02589737  5.2023 2.079e-07 ***
factor(year)2003         0.27753556  0.02578058 10.7653 < 2.2e-16 ***
factor(year)2004         0.42083450  0.02583218 16.2911 < 2.2e-16 ***
factor(year)2005         0.54635905  0.02581778 21.1621 < 2.2e-16 ***
factor(year)2006         0.67803943  0.02584215 26.2377 < 2.2e-16 ***
factor(year)2007         0.83620998  0.02591677 32.2652 < 2.2e-16 ***
factor(year)2008         0.98494465  0.02601334 37.8631 < 2.2e-16 ***
factor(year)2009         0.92998253  0.02598363 35.7911 < 2.2e-16 ***
factor(year)2010         1.02565309  0.02599646 39.4536 < 2.2e-16 ***
factor(year)2011         1.14098512  0.02608085 43.7480 < 2.2e-16 ***
factor(year)2012         1.17980726  0.02620535 45.0216 < 2.2e-16 ***
factor(year)2013         1.22693690  0.02620088 46.8281 < 2.2e-16 ***
factor(year)2014         1.26221752  0.02633785 47.9241 < 2.2e-16 ***
factor(year)2015         1.18450850  0.02642152 44.8312 < 2.2e-16 ***
factor(year)2016         1.19478238  0.02662248 44.8787 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1093.7
Residual Sum of Squares: 224.99
R-Squared:      0.79428
Adj. R-Squared: 0.78168
F-statistic: 693.611 on 20 and 3593 DF, p-value: < 2.22e-16
```

Model 5

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ specialization + related.diversification +
    factor(year) + factor(country), data = panel.data, model = "within",
    index = c("country", "year"))

Unbalanced Panel: n = 201, T = 1-20, N = 3814

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.0871705 -0.1365496  0.0065542  0.1435242  0.9826689

Coefficients:
                          Estimate Std. Error t-value  Pr(>|t|)
specialization           0.0899068  0.0398318  2.2572   0.02406 *
related.diversification -0.1772448  0.0411874 -4.3034 1.727e-05 ***
factor(year)1998        -0.0079232  0.0260476 -0.3042   0.76101
factor(year)1999         0.0066622  0.0261340  0.2549   0.79880
factor(year)2000         0.0509857  0.0259317  1.9662   0.04936 *
factor(year)2001         0.0588471  0.0258535  2.2762   0.02289 *
factor(year)2002         0.1334129  0.0257957  5.1719 2.444e-07 ***
factor(year)2003         0.2716421  0.0256984 10.5704 < 2.2e-16 ***
factor(year)2004         0.4151120  0.0257509 16.1203 < 2.2e-16 ***
factor(year)2005         0.5417756  0.0257315 21.0549 < 2.2e-16 ***
factor(year)2006         0.6718794  0.0257692 26.0729 < 2.2e-16 ***
factor(year)2007         0.8304874  0.0258537 32.1226 < 2.2e-16 ***
factor(year)2008         0.9781061  0.0259734 37.6580 < 2.2e-16 ***
factor(year)2009         0.9212552  0.0259302 35.5283 < 2.2e-16 ***
factor(year)2010         1.0167920  0.0259527 39.1787 < 2.2e-16 ***
factor(year)2011         1.1319255  0.0260288 43.4874 < 2.2e-16 ***
factor(year)2012         1.1692741  0.0262025 44.6246 < 2.2e-16 ***
factor(year)2013         1.2166550  0.0261868 46.4606 < 2.2e-16 ***
factor(year)2014         1.2519829  0.0263348 47.5410 < 2.2e-16 ***
factor(year)2015         1.1726523  0.0264541 44.3279 < 2.2e-16 ***
factor(year)2016         1.1840221  0.0266411 44.4434 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1093.7
Residual Sum of Squares: 223.12
R-Squared:      0.79599
Adj. R-Squared: 0.78344
F-statistic: 667.381 on 21 and 3592 DF, p-value: < 2.22e-16
```

Model 6
Oneway (individual) effect Within Model

```
Call:
plm(formula = logGDP ~ specialization + unrelated.diversification +
    factor(year) + factor(country), data = panel.data, model = "within",
    index = c("country", "year"))

Unbalanced Panel: n = 201, T = 1-20, N = 3814

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.0871705 -0.1365496  0.0065542  0.1435242  0.9826689

Coefficients:
                            Estimate Std. Error t-value  Pr(>|t|)
specialization             0.2671516  0.0430046  6.2122 5.823e-10 ***
unrelated.diversification  0.1772448  0.0411874  4.3034 1.727e-05 ***
factor(year)1998          -0.0079232  0.0260476 -0.3042   0.76101
factor(year)1999           0.0066622  0.0261340  0.2549   0.79880
factor(year)2000           0.0509857  0.0259317  1.9662   0.04936 *
factor(year)2001           0.0588471  0.0258535  2.2762   0.02289 *
factor(year)2002           0.1334129  0.0257957  5.1719 2.444e-07 ***
factor(year)2003           0.2716421  0.0256984 10.5704 < 2.2e-16 ***
factor(year)2004           0.4151120  0.0257509 16.1203 < 2.2e-16 ***
factor(year)2005           0.5417756  0.0257315 21.0549 < 2.2e-16 ***
factor(year)2006           0.6718794  0.0257692 26.0729 < 2.2e-16 ***
factor(year)2007           0.8304874  0.0258537 32.1226 < 2.2e-16 ***
factor(year)2008           0.9781061  0.0259734 37.6580 < 2.2e-16 ***
factor(year)2009           0.9212552  0.0259302 35.5283 < 2.2e-16 ***
factor(year)2010           1.0167920  0.0259527 39.1787 < 2.2e-16 ***
factor(year)2011           1.1319255  0.0260288 43.4874 < 2.2e-16 ***
factor(year)2012           1.1692741  0.0262025 44.6246 < 2.2e-16 ***
factor(year)2013           1.2166550  0.0261868 46.4606 < 2.2e-16 ***
factor(year)2014           1.2519829  0.0263348 47.5410 < 2.2e-16 ***
factor(year)2015           1.1726523  0.0264541 44.3279 < 2.2e-16 ***
factor(year)2016           1.1840221  0.0266411 44.4434 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1093.7
Residual Sum of Squares: 223.12
R-Squared:      0.79599
Adj. R-Squared: 0.78344
F-statistic: 667.381 on 21 and 3592 DF, p-value: < 2.22e-16
```

Model 7

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ specialization + factor(year) + factor(country),
   data = panel.data, model = "within", index = c("country",
      "year"))

Unbalanced Panel: n = 239, T = 1-20, N = 4471

Residuals:
    Min.   1st Qu.    Median   3rd Qu.      Max.
-3.925791 -0.203502  0.004014  0.257155  2.038025

Coefficients:
                 Estimate Std. Error t-value  Pr(>|t|)
specialization    0.2202152  0.0605009  3.6399 0.0002761 ***
factor(year)1998  0.0015945  0.0494970  0.0322 0.9743024
factor(year)1999  0.0182333  0.0494589  0.3687 0.7124025
factor(year)2000 -0.0445770  0.0493867 -0.9026 0.3667834
factor(year)2001  0.0473918  0.0492938  0.9614 0.3363986
factor(year)2002  0.1529378  0.0492790  3.1035 0.0019251 **
factor(year)2003  0.1886853  0.0491963  3.8354 0.0001272 ***
factor(year)2004  0.2917579  0.0492081  5.9291 3.291e-09 ***
factor(year)2005  0.3874824  0.0492180  7.8728 4.379e-15 ***
factor(year)2006  0.5613086  0.0491621 11.4175 < 2.2e-16 ***
factor(year)2007  0.6144249  0.0492044 12.4872 < 2.2e-16 ***
factor(year)2008  0.5973550  0.0494489 12.0802 < 2.2e-16 ***
factor(year)2009  0.7552116  0.0490256 15.4044 < 2.2e-16 ***
factor(year)2010  0.7569478  0.0491286 15.4075 < 2.2e-16 ***
factor(year)2011  0.8175486  0.0492838 16.5886 < 2.2e-16 ***
factor(year)2012  0.8583095  0.0494125 17.3703 < 2.2e-16 ***
factor(year)2013  0.9167814  0.0493826 18.5649 < 2.2e-16 ***
factor(year)2014  0.9737896  0.0494431 19.6952 < 2.2e-16 ***
factor(year)2015  0.9930707  0.0495353 20.0477 < 2.2e-16 ***
factor(year)2016  1.0241724  0.0494824 20.6977 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1768.7
Residual Sum of Squares: 1109.7
R-Squared:      0.37258
Adj. R-Squared: 0.33415
F-statistic: 125.061 on 20 and 4212 DF, p-value: < 2.22e-16
```

Model 8

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ `rel.div. (% amounts)` + factor(year) +
    factor(country), data = panel.data, model = "within", index = c("country",
    "year"))

Unbalanced Panel: n = 239, T = 1-20, N = 4471

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-3.7638169 -0.2076021  0.0063816  0.2665039  1.9081843

Coefficients:
                        Estimate Std. Error t-value  Pr(>|t|)
`rel.div. (% amounts)`  0.0074665  0.0015724  4.7484 2.119e-06 ***
factor(year)1998        0.0179621  0.0494369  0.3633  0.716373
factor(year)1999        0.0297786  0.0493890  0.6029  0.546581
factor(year)2000       -0.0304981  0.0493679 -0.6178  0.536759
factor(year)2001        0.0538808  0.0492338  1.0944  0.273847
factor(year)2002        0.1606188  0.0492397  3.2620  0.001115 **
factor(year)2003        0.2009985  0.0491332  4.0909 4.378e-05 ***
factor(year)2004        0.3135548  0.0491045  6.3855 1.894e-10 ***
factor(year)2005        0.4117605  0.0490383  8.3967 < 2.2e-16 ***
factor(year)2006        0.5866171  0.0489858 11.9753 < 2.2e-16 ***
factor(year)2007        0.6481483  0.0488568 13.2663 < 2.2e-16 ***
factor(year)2008        0.6350725  0.0490616 12.9444 < 2.2e-16 ***
factor(year)2009        0.7838000  0.0486911 16.0974 < 2.2e-16 ***
factor(year)2010        0.7938936  0.0486394 16.3220 < 2.2e-16 ***
factor(year)2011        0.8522559  0.0488098 17.4607 < 2.2e-16 ***
factor(year)2012        0.9008368  0.0489370 18.4081 < 2.2e-16 ***
factor(year)2013        0.9619514  0.0487745 19.7224 < 2.2e-16 ***
factor(year)2014        1.0178127  0.0486463 20.9227 < 2.2e-16 ***
factor(year)2015        1.0423325  0.0487446 21.3835 < 2.2e-16 ***
factor(year)2016        1.0750146  0.0486722 22.0868 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1768.7
Residual Sum of Squares: 1107.3
R-Squared:      0.37396
Adj. R-Squared: 0.33561
F-statistic: 125.8 on 20 and 4212 DF, p-value: < 2.22e-16
```

Model 9

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ `unrel.div (% amounts)` + factor(year) +
    factor(country), data = panel.data, model = "within", index = c("country",
    "year"))

Unbalanced Panel: n = 239, T = 1-20, N = 4471

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-3.9682779 -0.2026147  0.0030157  0.2540598  2.0133950

Coefficients:
                          Estimate  Std. Error t-value  Pr(>|t|)
`unrel.div (% amounts)` -0.00330396  0.00060372 -5.4727 4.690e-08 ***
factor(year)1998         0.00185899  0.04937494  0.0377 0.9699681
factor(year)1999         0.01755685  0.04934841  0.3558 0.7220281
factor(year)2000        -0.04198899  0.04927853 -0.8521 0.3942211
factor(year)2001         0.04689939  0.04919252  0.9534 0.3404500
factor(year)2002         0.15483739  0.04918004  3.1484 0.0016533 **
factor(year)2003         0.18851464  0.04908491  3.8406 0.0001245 ***
factor(year)2004         0.29069792  0.04906486  5.9248 3.377e-09 ***
factor(year)2005         0.38432299  0.04906357  7.8332 5.979e-15 ***
factor(year)2006         0.55843281  0.04900234 11.3960 < 2.2e-16 ***
factor(year)2007         0.60895052  0.04899887 12.4278 < 2.2e-16 ***
factor(year)2008         0.59185818  0.04921453 12.0261 < 2.2e-16 ***
factor(year)2009         0.74886803  0.04885651 15.3279 < 2.2e-16 ***
factor(year)2010         0.74937753  0.04890758 15.3223 < 2.2e-16 ***
factor(year)2011         0.80969022  0.04908124 16.4969 < 2.2e-16 ***
factor(year)2012         0.85215852  0.04914070 17.3412 < 2.2e-16 ***
factor(year)2013         0.90900727  0.04909283 18.5161 < 2.2e-16 ***
factor(year)2014         0.96276689  0.04918451 19.5746 < 2.2e-16 ***
factor(year)2015         0.98345706  0.04921554 19.9827 < 2.2e-16 ***
factor(year)2016         1.01466162  0.04914550 20.6461 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1768.7
Residual Sum of Squares: 1105.3
R-Squared:      0.37505
Adj. R-Squared: 0.33677
F-statistic: 126.388 on 20 and 4212 DF, p-value: < 2.22e-16
```

Model 10

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ specialization + related.diversification +
   factor(year) + factor(country), data = panel.data, model = "within",
   index = c("country", "year"))

Unbalanced Panel: n = 239, T = 1-20, N = 4471

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-3.8970076 -0.2024094  0.0039768  0.2581814  2.0493283

Coefficients:
                        Estimate Std. Error t-value  Pr(>|t|)
specialization          0.2844632  0.0646514  4.4000 1.110e-05 ***
related.diversification  0.1891029  0.0674651  2.8030  0.005087 **
factor(year)1998         0.0021323  0.0494571  0.0431  0.965613
factor(year)1999         0.0208515  0.0494275  0.4219  0.673148
factor(year)2000        -0.0416054  0.0493579 -0.8429  0.399314
factor(year)2001         0.0538704  0.0493079  1.0925  0.274663
factor(year)2002         0.1525689  0.0492392  3.0985  0.001958 **
factor(year)2003         0.1931945  0.0491826  3.9281 8.699e-05 ***
factor(year)2004         0.2951814  0.0491833  6.0017 2.117e-09 ***
factor(year)2005         0.3900094  0.0491863  7.9292 2.802e-15 ***
factor(year)2006         0.5655795  0.0491458 11.5082 < 2.2e-16 ***
factor(year)2007         0.6165108  0.0491700 12.5383 < 2.2e-16 ***
factor(year)2008         0.5979840  0.0494092 12.1027 < 2.2e-16 ***
factor(year)2009         0.7581328  0.0489968 15.4731 < 2.2e-16 ***
factor(year)2010         0.7597172  0.0490986 15.4733 < 2.2e-16 ***
factor(year)2011         0.8202113  0.0492529 16.6531 < 2.2e-16 ***
factor(year)2012         0.8603911  0.0493779 17.4246 < 2.2e-16 ***
factor(year)2013         0.9193104  0.0493507 18.6281 < 2.2e-16 ***
factor(year)2014         0.9753071  0.0494058 19.7407 < 2.2e-16 ***
factor(year)2015         0.9958288  0.0495048 20.1158 < 2.2e-16 ***
factor(year)2016         1.0254086  0.0494442 20.7387 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1768.7
Residual Sum of Squares: 1107.6
R-Squared:      0.37375
Adj. R-Squared: 0.33523
F-statistic: 119.674 on 21 and 4211 DF, p-value: < 2.22e-16
```

Model 11

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ specialization + unrelated.diversification +
   factor(year) + factor(country), data = panel.data, model = "within",
   index = c("country", "year"))

Unbalanced Panel: n = 239, T = 1-20, N = 4471

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-3.8970076 -0.2024094  0.0039768  0.2581814  2.0493283

Coefficients:
                          Estimate Std. Error t-value  Pr(>|t|)
specialization             0.0953603  0.0750904  1.2699  0.204176
unrelated.diversification -0.1891029  0.0674651 -2.8030  0.005087 **
factor(year)1998           0.0021323  0.0494571  0.0431  0.965613
factor(year)1999           0.0208515  0.0494275  0.4219  0.673148
factor(year)2000          -0.0416054  0.0493579 -0.8429  0.399314
factor(year)2001           0.0538704  0.0493079  1.0925  0.274663
factor(year)2002           0.1525689  0.0492392  3.0985  0.001958 **
factor(year)2003           0.1931945  0.0491826  3.9281 8.699e-05 ***
factor(year)2004           0.2951814  0.0491833  6.0017 2.117e-09 ***
factor(year)2005           0.3900094  0.0491863  7.9292 2.802e-15 ***
factor(year)2006           0.5655795  0.0491458 11.5082 < 2.2e-16 ***
factor(year)2007           0.6165108  0.0491700 12.5383 < 2.2e-16 ***
factor(year)2008           0.5979840  0.0494092 12.1027 < 2.2e-16 ***
factor(year)2009           0.7581328  0.0489968 15.4731 < 2.2e-16 ***
factor(year)2010           0.7597172  0.0490986 15.4733 < 2.2e-16 ***
factor(year)2011           0.8202113  0.0492529 16.6531 < 2.2e-16 ***
factor(year)2012           0.8603911  0.0493779 17.4246 < 2.2e-16 ***
factor(year)2013           0.9193104  0.0493507 18.6281 < 2.2e-16 ***
factor(year)2014           0.9753071  0.0494058 19.7407 < 2.2e-16 ***
factor(year)2015           0.9958288  0.0495048 20.1158 < 2.2e-16 ***
factor(year)2016           1.0254086  0.0494442 20.7387 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1768.7
Residual Sum of Squares: 1107.6
R-Squared:      0.37375
Adj. R-Squared: 0.33523
F-statistic: 119.674 on 21 and 4211 DF, p-value: < 2.22e-16
```

Model 12

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ ubiquity + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))

Unbalanced Panel: n = 239, T = 1-21, N = 4684

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-3.617572 -0.210451  0.015935  0.267141  1.946048

Coefficients:
                   Estimate  Std. Error  t-value  Pr(>|t|)
ubiquity        -0.01202397  0.00095888 -12.5396 < 2.2e-16 ***
factor(year)1997  0.03541747  0.04918209   0.7201   0.47148
factor(year)1998  0.06780573  0.04931140   1.3751   0.16919
factor(year)1999  0.07736551  0.04925953   1.5706   0.11635
factor(year)2000  0.00325953  0.04922754   0.0662   0.94721
factor(year)2001  0.09185867  0.04913693   1.8694   0.06163 .
factor(year)2002  0.20302125  0.04914837   4.1308 3.682e-05 ***
factor(year)2003  0.28101473  0.04910366   5.7229 1.116e-08 ***
factor(year)2004  0.39203249  0.04909063   7.9859 1.764e-15 ***
factor(year)2005  0.53029926  0.04934731  10.7463 < 2.2e-16 ***
factor(year)2006  0.72679859  0.04955770  14.6657 < 2.2e-16 ***
factor(year)2007  0.77002161  0.04920312  15.6499 < 2.2e-16 ***
factor(year)2008  0.76573638  0.04950694  15.4673 < 2.2e-16 ***
factor(year)2009  0.94795437  0.04957343  19.1222 < 2.2e-16 ***
factor(year)2010  0.97302563  0.04980379  19.5372 < 2.2e-16 ***
factor(year)2011  1.04410278  0.05025329  20.7768 < 2.2e-16 ***
factor(year)2012  1.09637004  0.05047195  21.7224 < 2.2e-16 ***
factor(year)2013  1.17088937  0.05057857  23.1499 < 2.2e-16 ***
factor(year)2014  1.24197964  0.05076153  24.4669 < 2.2e-16 ***
factor(year)2015  1.25709643  0.05070183  24.7939 < 2.2e-16 ***
factor(year)2016  1.29488143  0.05077328  25.5032 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1908.5
Residual Sum of Squares: 1147.2
R-Squared:      0.39889
Adj. R-Squared: 0.3637
F-statistic: 139.797 on 21 and 4424 DF, p-value: < 2.22e-16
```

Model 13

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ nr.of.fields + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))

Unbalanced Panel: n = 239, T = 1-21, N = 4684

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-3.6270894 -0.2014390  0.0097965  0.2490599  2.0719206

Coefficients:
                 Estimate  Std. Error t-value  Pr(>|t|)
nr.of.fields      0.00597758  0.00039423 15.1626 < 2.2e-16 ***
factor(year)1997  0.02904833  0.04880235  0.5952  0.551724
factor(year)1998  0.03084968  0.04894053  0.6304  0.528498
factor(year)1999  0.04260782  0.04890426  0.8712  0.383665
factor(year)2000 -0.03447261  0.04891310 -0.7048  0.480989
factor(year)2001  0.04670759  0.04886850  0.9558  0.339235
factor(year)2002  0.12743986  0.04900687  2.6004  0.009341 **
factor(year)2003  0.14635877  0.04902204  2.9856  0.002846 **
factor(year)2004  0.22859506  0.04924316  4.6422 3.548e-06 ***
factor(year)2005  0.28922904  0.04966138  5.8240 6.152e-09 ***
factor(year)2006  0.42617234  0.05019557  8.4902 < 2.2e-16 ***
factor(year)2007  0.46636645  0.05039436  9.2543 < 2.2e-16 ***
factor(year)2008  0.42647153  0.05105245  8.3536 < 2.2e-16 ***
factor(year)2009  0.55315442  0.05134839 10.7726 < 2.2e-16 ***
factor(year)2010  0.54831789  0.05153728 10.6392 < 2.2e-16 ***
factor(year)2011  0.58600096  0.05227138 11.2107 < 2.2e-16 ***
factor(year)2012  0.60521650  0.05290530 11.4396 < 2.2e-16 ***
factor(year)2013  0.65065043  0.05323128 12.2231 < 2.2e-16 ***
factor(year)2014  0.69080213  0.05372985 12.8570 < 2.2e-16 ***
factor(year)2015  0.69600077  0.05419203 12.8432 < 2.2e-16 ***
factor(year)2016  0.71276617  0.05458761 13.0573 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1908.5
Residual Sum of Squares: 1129.3
R-Squared:      0.40828
Adj. R-Squared: 0.37364
F-statistic: 145.356 on 21 and 4424 DF, p-value: < 2.22e-16
```

Model 14

```
Oneway (individual) effect Within Model

Call:
plm(formula = entropy ~ hoover.specialization + factor(year) +
    factor(country), data = panel.data, model = "within", index = c("country",
    "year"))

Unbalanced Panel: n = 239, T = 1-21, N = 4684

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-3.1120840 -0.1397037  0.0038752  0.1704444  1.9132501

Coefficients:
                        Estimate Std. Error  t-value  Pr(>|t|)
hoover.specialization -5.8940378  0.1262983 -46.6676 < 2.2e-16 ***
factor(year)1997      -0.0097500  0.0409809  -0.2379   0.81196
factor(year)1998       0.0049575  0.0410829   0.1207   0.90396
factor(year)1999       0.0330665  0.0410395   0.8057   0.42045
factor(year)2000      -0.0518075  0.0410260  -1.2628   0.20673
factor(year)2001       0.0042525  0.0409797   0.1038   0.91736
factor(year)2002       0.0679160  0.0410416   1.6548   0.09803 .
factor(year)2003       0.0633811  0.0409985   1.5459   0.12219
factor(year)2004       0.0747148  0.0412205   1.8126   0.06997 .
factor(year)2005       0.1027180  0.0414185   2.4800   0.01318 *
factor(year)2006       0.2140457  0.0416402   5.1404 2.860e-07 ***
factor(year)2007       0.2298549  0.0417220   5.5092 3.809e-08 ***
factor(year)2008       0.1766544  0.0420640   4.1997 2.726e-05 ***
factor(year)2009       0.2699668  0.0421459   6.4055 1.655e-10 ***
factor(year)2010       0.2698695  0.0421178   6.4075 1.634e-10 ***
factor(year)2011       0.2764802  0.0426326   6.4852 9.834e-11 ***
factor(year)2012       0.2504743  0.0431285   5.8076 6.780e-09 ***
factor(year)2013       0.2674523  0.0433455   6.1702 7.428e-10 ***
factor(year)2014       0.2903811  0.0435770   6.6636 2.999e-11 ***
factor(year)2015       0.2733959  0.0439086   6.2265 5.213e-10 ***
factor(year)2016       0.2633643  0.0442094   5.9572 2.765e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1908.5
Residual Sum of Squares: 796.09
R-Squared:      0.58287
Adj. R-Squared: 0.55845
F-statistic: 294.375 on 21 and 4424 DF, p-value: < 2.22e-16
```

Model 15

Oneway (individual) effect Within Model

Call:
```
plm(formula = entropy ~ publication.count + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
      "year"))
```

Unbalanced Panel: n = 239, T = 1-21, N = 4684

Residuals:
```
     Min.   1st Qu.    Median   3rd Qu.      Max.
-3.682617 -0.209311  0.010023  0.258519  2.045796
```

Coefficients:
```
                   Estimate  Std. Error t-value  Pr(>|t|)
publication.count -1.8525e-06  2.2015e-07 -8.4146 < 2.2e-16 ***
factor(year)1997   4.8407e-02  4.9645e-02  0.9751   0.32958
factor(year)1998   5.6729e-02  4.9773e-02  1.1398   0.25445
factor(year)1999   7.2687e-02  4.9728e-02  1.4617   0.14390
factor(year)2000   9.4588e-03  4.9700e-02  0.1903   0.84907
factor(year)2001   1.0329e-01  4.9608e-02  2.0821   0.03739 *
factor(year)2002   2.0764e-01  4.9624e-02  4.1843 2.916e-05 ***
factor(year)2003   2.5050e-01  4.9483e-02  5.0624 4.308e-07 ***
factor(year)2004   3.6143e-01  4.9463e-02  7.3070 3.224e-13 ***
factor(year)2005   4.6459e-01  4.9421e-02  9.4006 < 2.2e-16 ***
factor(year)2006   6.4245e-01  4.9386e-02 13.0089 < 2.2e-16 ***
factor(year)2007   7.0482e-01  4.9245e-02 14.3124 < 2.2e-16 ***
factor(year)2008   6.9276e-01  4.9445e-02 14.0109 < 2.2e-16 ***
factor(year)2009   8.5137e-01  4.9139e-02 17.3257 < 2.2e-16 ***
factor(year)2010   8.6044e-01  4.9071e-02 17.5345 < 2.2e-16 ***
factor(year)2011   9.2339e-01  4.9337e-02 18.7158 < 2.2e-16 ***
factor(year)2012   9.6831e-01  4.9386e-02 19.6069 < 2.2e-16 ***
factor(year)2013   1.0333e+00  4.9255e-02 20.9792 < 2.2e-16 ***
factor(year)2014   1.0974e+00  4.9228e-02 22.2927 < 2.2e-16 ***
factor(year)2015   1.1171e+00  4.9274e-02 22.6711 < 2.2e-16 ***
factor(year)2016   1.1511e+00  4.9229e-02 23.3818 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Total Sum of Squares: 1908.5
Residual Sum of Squares: 1169.3
R-Squared: 0.38733
Adj. R-Squared: 0.35146
F-statistic: 133.185 on 21 and 4424 DF, p-value: < 2.22e-16

Model 16
Oneway (individual) effect Within Model

```
Call:
plm(formula = entropy ~ citations + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
      "year"))

Unbalanced Panel: n = 239, T = 1-21, N = 4684

Residuals:
    Min.   1st Qu.    Median   3rd Qu.      Max.
-3.702584 -0.213852  0.010805  0.267994  2.026857

Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)
citations        2.8528e-08 9.8064e-09  2.9091  0.003643 **
factor(year)1997 4.5910e-02 4.9993e-02  0.9183  0.358495
factor(year)1998 5.3428e-02 5.0125e-02  1.0659  0.286525
factor(year)1999 6.8207e-02 5.0081e-02  1.3619  0.173292
factor(year)2000 2.5032e-03 5.0058e-02  0.0500  0.960121
factor(year)2001 9.5235e-02 4.9964e-02  1.9061  0.056708 .
factor(year)2002 1.9768e-01 4.9983e-02  3.9548 7.779e-05 ***
factor(year)2003 2.3657e-01 4.9847e-02  4.7459 2.141e-06 ***
factor(year)2004 3.4430e-01 4.9825e-02  6.9101 5.532e-12 ***
factor(year)2005 4.4302e-01 4.9771e-02  8.9012 < 2.2e-16 ***
factor(year)2006 6.1750e-01 4.9719e-02 12.4197 < 2.2e-16 ***
factor(year)2007 6.7714e-01 4.9561e-02 13.6628 < 2.2e-16 ***
factor(year)2008 6.6244e-01 4.9738e-02 13.3186 < 2.2e-16 ***
factor(year)2009 8.1751e-01 4.9395e-02 16.5504 < 2.2e-16 ***
factor(year)2010 8.2504e-01 4.9298e-02 16.7359 < 2.2e-16 ***
factor(year)2011 8.8522e-01 4.9520e-02 17.8760 < 2.2e-16 ***
factor(year)2012 9.2935e-01 4.9537e-02 18.7605 < 2.2e-16 ***
factor(year)2013 9.9383e-01 4.9373e-02 20.1290 < 2.2e-16 ***
factor(year)2014 1.0560e+00 4.9302e-02 21.4191 < 2.2e-16 ***
factor(year)2015 1.0781e+00 4.9369e-02 21.8380 < 2.2e-16 ***
factor(year)2016 1.1123e+00 4.9332e-02 22.5479 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1908.5
Residual Sum of Squares: 1185.7
R-Squared:      0.37872
Adj. R-Squared: 0.34234
F-statistic: 128.416 on 21 and 4424 DF, p-value: < 2.22e-16
```

Model 17

Oneway (individual) effect Within Model

```
Call:
plm(formula = entropy ~ `H index` + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))

Unbalanced Panel: n = 239, T = 1-21, N = 4684

Residuals:
    Min.   1st Qu.    Median   3rd Qu.      Max.
-3.720301 -0.211918  0.011977  0.267600  2.046394

Coefficients:
                  Estimate  Std. Error t-value  Pr(>|t|)
`H index`         2.0006e-04  3.6281e-05  5.5142 3.703e-08 ***
factor(year)1997  3.8694e-02  4.9890e-02  0.7756  0.438039
factor(year)1998  4.2813e-02  5.0047e-02  0.8554  0.392350
factor(year)1999  5.4258e-02  5.0037e-02  1.0844  0.278264
factor(year)2000 -1.7559e-02  5.0101e-02 -0.3505  0.725998
factor(year)2001  7.1348e-02  5.0071e-02  1.4249  0.154249
factor(year)2002  1.6546e-01  5.0269e-02  3.2914  0.001005 **
factor(year)2003  1.9659e-01  5.0359e-02  3.9038 9.611e-05 ***
factor(year)2004  2.9690e-01  5.0582e-02  5.8697 4.685e-09 ***
factor(year)2005  3.8495e-01  5.0941e-02  7.5568 4.990e-14 ***
factor(year)2006  5.4997e-01  5.1310e-02 10.7186 < 2.2e-16 ***
factor(year)2007  6.0309e-01  5.1476e-02 11.7159 < 2.2e-16 ***
factor(year)2008  5.8139e-01  5.2004e-02 11.1796 < 2.2e-16 ***
factor(year)2009  7.3225e-01  5.1898e-02 14.1094 < 2.2e-16 ***
factor(year)2010  7.3617e-01  5.1978e-02 14.1630 < 2.2e-16 ***
factor(year)2011  7.9162e-01  5.2416e-02 15.1028 < 2.2e-16 ***
factor(year)2012  8.2964e-01  5.2731e-02 15.7335 < 2.2e-16 ***
factor(year)2013  8.9096e-01  5.2658e-02 16.9198 < 2.2e-16 ***
factor(year)2014  9.4855e-01  5.2735e-02 17.9872 < 2.2e-16 ***
factor(year)2015  9.6686e-01  5.2849e-02 18.2948 < 2.2e-16 ***
factor(year)2016  9.9765e-01  5.2842e-02 18.8797 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1908.5
Residual Sum of Squares: 1179.9
R-Squared:      0.38178
Adj. R-Squared: 0.34558
F-statistic: 130.095 on 21 and 4424 DF, p-value: < 2.22e-16
```

Model 18

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ ubiquity + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))

Unbalanced Panel: n = 201, T = 2-21, N = 3998

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.3284008 -0.1435835  0.0060078  0.1451977  1.0548218

Coefficients:
                  Estimate  Std. Error t-value  Pr(>|t|)
ubiquity         -0.00339939  0.00066363 -5.1224 3.167e-07 ***
factor(year)1997  0.03057667  0.02670857  1.1448 0.2523537
factor(year)1998  0.03178907  0.02673238  1.1892 0.2344515
factor(year)1999  0.04948795  0.02681858  1.8453 0.0650742 .
factor(year)2000  0.08783277  0.02663548  3.2976 0.0009842 ***
factor(year)2001  0.09963270  0.02651803  3.7572 0.0001744 ***
factor(year)2002  0.17409787  0.02649786  6.5703 5.709e-11 ***
factor(year)2003  0.32420773  0.02641046 12.2757 < 2.2e-16 ***
factor(year)2004  0.47113947  0.02647385 17.7964 < 2.2e-16 ***
factor(year)2005  0.60757131  0.02667208 22.7793 < 2.2e-16 ***
factor(year)2006  0.74972106  0.02692564 27.8441 < 2.2e-16 ***
factor(year)2007  0.90742773  0.02683216 33.8187 < 2.2e-16 ***
factor(year)2008  1.05864050  0.02703646 39.1560 < 2.2e-16 ***
factor(year)2009  1.00951578  0.02728341 37.0011 < 2.2e-16 ***
factor(year)2010  1.11192013  0.02752756 40.3930 < 2.2e-16 ***
factor(year)2011  1.22838138  0.02762191 44.4713 < 2.2e-16 ***
factor(year)2012  1.27270552  0.02802128 45.4193 < 2.2e-16 ***
factor(year)2013  1.32410116  0.02811535 47.0953 < 2.2e-16 ***
factor(year)2014  1.36420204  0.02842526 47.9926 < 2.2e-16 ***
factor(year)2015  1.28580122  0.02836006 45.3385 < 2.2e-16 ***
factor(year)2016  1.29498127  0.02848380 45.4638 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1201.1
Residual Sum of Squares: 247.06
R-Squared:      0.79431
Adj. R-Squared: 0.78227
F-statistic: 694.369 on 21 and 3776 DF, p-value: < 2.22e-16
```

Model 19
Oneway (individual) effect Within Model

```
Call:
plm(formula = logGDP ~ nr.of.fields + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))

Unbalanced Panel: n = 201, T = 2-21, N = 4039

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max.
-2.42955 -0.13258  0.00519  0.13704  1.11832

Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)
nr.of.fields     0.00382575 0.00020884 18.3187 < 2.2e-16 ***
factor(year)1997 0.02380810 0.02553747  0.9323  0.351250
factor(year)1998 0.01661122 0.02551158  0.6511  0.515005
factor(year)1999 0.03358038 0.02548250  1.3178  0.187656
factor(year)2000 0.05771260 0.02541475  2.2708  0.023213 *
factor(year)2001 0.06633106 0.02546044  2.6053  0.009216 **
factor(year)2002 0.12219732 0.02542670  4.8059   1.6e-06 ***
factor(year)2003 0.24511219 0.02553642  9.5985 < 2.2e-16 ***
factor(year)2004 0.37656400 0.02562594 14.6946 < 2.2e-16 ***
factor(year)2005 0.47778306 0.02590398 18.4444 < 2.2e-16 ***
factor(year)2006 0.58527483 0.02619619 22.3420 < 2.2e-16 ***
factor(year)2007 0.72846066 0.02642038 27.5719 < 2.2e-16 ***
factor(year)2008 0.86369754 0.02673517 32.3057 < 2.2e-16 ***
factor(year)2009 0.79192449 0.02710585 29.2160 < 2.2e-16 ***
factor(year)2010 0.87602040 0.02731571 32.0702 < 2.2e-16 ***
factor(year)2011 0.97872789 0.02764088 35.4087 < 2.2e-16 ***
factor(year)2012 1.00503216 0.02809331 35.7748 < 2.2e-16 ***
factor(year)2013 1.03880850 0.02839380 36.5857 < 2.2e-16 ***
factor(year)2014 1.05976989 0.02884753 36.7369 < 2.2e-16 ***
factor(year)2015 0.97323144 0.02912822 33.4120 < 2.2e-16 ***
factor(year)2016 0.97134061 0.02963857 32.7729 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1212.9
Residual Sum of Squares: 231.72
R-Squared:      0.80894
Adj. R-Squared: 0.79788
F-statistic: 769.596 on 21 and 3817 DF, p-value: < 2.22e-16
```

Model 20

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ hoover.specialization + factor(year) +
   factor(country), data = panel.data, model = "within", index = c("country",
   "year"))

Unbalanced Panel: n = 201, T = 2-21, N = 3998

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.3849720 -0.1434146  0.0040677  0.1446723  1.0730283

Coefficients:
                   Estimate Std. Error  t-value  Pr(>|t|)
hoover.specialization -0.906832   0.082655 -10.9713 < 2.2e-16 ***
factor(year)1997       0.027341   0.026380   1.0364 0.3000600
factor(year)1998       0.024707   0.026413   0.9354 0.3496397
factor(year)1999       0.041796   0.026491   1.5777 0.1147092
factor(year)2000       0.080875   0.026321   3.0727 0.0021367 **
factor(year)2001       0.088258   0.026222   3.3658 0.0007707 ***
factor(year)2002       0.150896   0.026237   5.7513 9.559e-09 ***
factor(year)2003       0.286125   0.026171  10.9331 < 2.2e-16 ***
factor(year)2004       0.417560   0.026361  15.8401 < 2.2e-16 ***
factor(year)2005       0.532741   0.026501  20.1026 < 2.2e-16 ***
factor(year)2006       0.656170   0.026631  24.6391 < 2.2e-16 ***
factor(year)2007       0.810945   0.026773  30.2898 < 2.2e-16 ***
factor(year)2008       0.953776   0.027004  35.3203 < 2.2e-16 ***
factor(year)2009       0.889659   0.027197  32.7114 < 2.2e-16 ***
factor(year)2010       0.983168   0.027250  36.0789 < 2.2e-16 ***
factor(year)2011       1.090384   0.027561  39.5620 < 2.2e-16 ***
factor(year)2012       1.119461   0.027972  40.0204 < 2.2e-16 ***
factor(year)2013       1.160612   0.028133  41.2552 < 2.2e-16 ***
factor(year)2014       1.189791   0.028454  41.8143 < 2.2e-16 ***
factor(year)2015       1.108552   0.028641  38.7050 < 2.2e-16 ***
factor(year)2016       1.110169   0.029163  38.0678 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1201.1
Residual Sum of Squares: 241.09
R-Squared:      0.79928
Adj. R-Squared: 0.78753
F-statistic: 716.01 on 21 and 3776 DF, p-value: < 2.22e-16
```

Model 21

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ publication.count + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
        "year"))

Unbalanced Panel: n = 201, T = 2-21, N = 3998

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.379605 -0.142334  0.004206  0.147812  1.041282

Coefficients:
                    Estimate  Std. Error t-value  Pr(>|t|)
publication.count -5.7352e-08  1.1067e-07 -0.5182 0.6043188
factor(year)1997   3.5008e-02  2.6787e-02  1.3069 0.1913191
factor(year)1998   3.0926e-02  2.6824e-02  1.1529 0.2490147
factor(year)1999   4.6829e-02  2.6906e-02  1.7405 0.0818535 .
factor(year)2000   8.8807e-02  2.6728e-02  3.3227 0.0009001 ***
factor(year)2001   1.0204e-01  2.6608e-02  3.8349 0.0001277 ***
factor(year)2002   1.7161e-01  2.6587e-02  6.4546 1.223e-10 ***
factor(year)2003   3.1574e-01  2.6453e-02 11.9359 < 2.2e-16 ***
factor(year)2004   4.6084e-01  2.6494e-02 17.3938 < 2.2e-16 ***
factor(year)2005   5.8732e-01  2.6474e-02 22.1847 < 2.2e-16 ***
factor(year)2006   7.2092e-01  2.6427e-02 27.2800 < 2.2e-16 ***
factor(year)2007   8.8160e-01  2.6458e-02 33.3211 < 2.2e-16 ***
factor(year)2008   1.0303e+00  2.6575e-02 38.7709 < 2.2e-16 ***
factor(year)2009   9.7530e-01  2.6566e-02 36.7120 < 2.2e-16 ***
factor(year)2010   1.0724e+00  2.6537e-02 40.4120 < 2.2e-16 ***
factor(year)2011   1.1885e+00  2.6629e-02 44.6324 < 2.2e-16 ***
factor(year)2012   1.2279e+00  2.6751e-02 45.9025 < 2.2e-16 ***
factor(year)2013   1.2766e+00  2.6679e-02 47.8516 < 2.2e-16 ***
factor(year)2014   1.3133e+00  2.6786e-02 49.0301 < 2.2e-16 ***
factor(year)2015   1.2364e+00  2.6825e-02 46.0911 < 2.2e-16 ***
factor(year)2016   1.2468e+00  2.7068e-02 46.0640 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1201.1
Residual Sum of Squares: 248.76
R-Squared:      0.7929
Adj. R-Squared: 0.78077
F-statistic: 688.398 on 21 and 3776 DF, p-value: < 2.22e-16
```

Model 22

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ citations + factor(year) + factor(country),
    data = panel.data, model = "within", index = c("country",
       "year"))

Unbalanced Panel: n = 201, T = 2-21, N = 3998

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.3833571 -0.1424818  0.0032742  0.1442932  1.0426583

Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)
citations        3.5573e-08 4.8421e-09  7.3467 2.477e-13 ***
factor(year)1997 3.3795e-02 2.6598e-02  1.2706 0.2039572
factor(year)1998 2.8578e-02 2.6637e-02  1.0729 0.2833947
factor(year)1999 4.3796e-02 2.6719e-02  1.6391 0.1012712
factor(year)2000 8.4107e-02 2.6546e-02  3.1684 0.0015450 **
factor(year)2001 9.7212e-02 2.6426e-02  3.6787 0.0002377 ***
factor(year)2002 1.6598e-01 2.6407e-02  6.2855 3.639e-10 ***
factor(year)2003 3.0842e-01 2.6276e-02 11.7378 < 2.2e-16 ***
factor(year)2004 4.5270e-01 2.6316e-02 17.2023 < 2.2e-16 ***
factor(year)2005 5.7863e-01 2.6288e-02 22.0113 < 2.2e-16 ***
factor(year)2006 7.1209e-01 2.6231e-02 27.1471 < 2.2e-16 ***
factor(year)2007 8.7266e-01 2.6252e-02 33.2412 < 2.2e-16 ***
factor(year)2008 1.0217e+00 2.6354e-02 38.7700 < 2.2e-16 ***
factor(year)2009 9.6699e-01 2.6325e-02 36.7332 < 2.2e-16 ***
factor(year)2010 1.0651e+00 2.6278e-02 40.5304 < 2.2e-16 ***
factor(year)2011 1.1826e+00 2.6341e-02 44.8959 < 2.2e-16 ***
factor(year)2012 1.2240e+00 2.6441e-02 46.2929 < 2.2e-16 ***
factor(year)2013 1.2756e+00 2.6348e-02 48.4127 < 2.2e-16 ***
factor(year)2014 1.3155e+00 2.6425e-02 49.7844 < 2.2e-16 ***
factor(year)2015 1.2426e+00 2.6477e-02 46.9305 < 2.2e-16 ***
factor(year)2016 1.2572e+00 2.6721e-02 47.0481 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1201.1
Residual Sum of Squares: 245.27
R-Squared:      0.7958
Adj. R-Squared: 0.78385
F-statistic: 700.746 on 21 and 3776 DF, p-value: < 2.22e-16
```

Model 23

```
Oneway (individual) effect Within Model

Call:
plm(formula = logGDP ~ `H index` + factor(year) + factor(country),
   data = panel.data, model = "within", index = c("country",
      "year"))

Unbalanced Panel: n = 201, T = 2-21, N = 3998

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-2.4497929 -0.1366198  0.0046521  0.1432662  1.0580975

Coefficients:
                 Estimate Std. Error t-value  Pr(>|t|)
`H index`        2.1082e-04 1.8697e-05 11.2760 < 2.2e-16 ***
factor(year)1997 2.6930e-02 2.6357e-02  1.0217  0.306967
factor(year)1998 1.8372e-02 2.6407e-02  0.6957  0.486645
factor(year)1999 3.0365e-02 2.6504e-02  1.1457  0.252009
factor(year)2000 6.2324e-02 2.6392e-02  2.3615  0.018253 *
factor(year)2001 6.9592e-02 2.6325e-02  2.6435  0.008239 **
factor(year)2002 1.2780e-01 2.6432e-02  4.8351 1.384e-06 ***
factor(year)2003 2.6076e-01 2.6459e-02  9.8553 < 2.2e-16 ***
factor(year)2004 3.9904e-01 2.6610e-02 14.9960 < 2.2e-16 ***
factor(year)2005 5.1302e-01 2.6826e-02 19.1237 < 2.2e-16 ***
factor(year)2006 6.3603e-01 2.7013e-02 23.5454 < 2.2e-16 ***
factor(year)2007 7.8897e-01 2.7229e-02 28.9759 < 2.2e-16 ***
factor(year)2008 9.2985e-01 2.7540e-02 33.7638 < 2.2e-16 ***
factor(year)2009 8.7016e-01 2.7646e-02 31.4755 < 2.2e-16 ***
factor(year)2010 9.6317e-01 2.7728e-02 34.7367 < 2.2e-16 ***
factor(year)2011 1.0744e+00 2.7942e-02 38.4501 < 2.2e-16 ***
factor(year)2012 1.1090e+00 2.8191e-02 39.3368 < 2.2e-16 ***
factor(year)2013 1.1558e+00 2.8165e-02 41.0363 < 2.2e-16 ***
factor(year)2014 1.1890e+00 2.8344e-02 41.9483 < 2.2e-16 ***
factor(year)2015 1.1107e+00 2.8422e-02 39.0797 < 2.2e-16 ***
factor(year)2016 1.1192e+00 2.8682e-02 39.0200 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1201.1
Residual Sum of Squares: 240.68
R-Squared:      0.79963
Adj. R-Squared: 0.7879
F-statistic: 717.569 on 21 and 3776 DF, p-value: < 2.22e-16
```

Model 24 - Type of economy (IV) vs. growth factor portfolio (DV) - between estimator model

```
Oneway (individual) effect Between Model

Call:
plm(formula = `growth factor` ~ X, data = panel.h4, model = "between")

Balanced Panel: n = 59, T = 21, N = 1239
Observations used in estimation: 59

Residuals:
      Min.     1st Qu.      Median     3rd Qu.        Max.
-0.19012744 -0.05080295 -0.00058164  0.04415053  0.23524354

Coefficients: (1 dropped because of singularities)
                     Estimate Std. Error t-value  Pr(>|t|)
(Intercept)          1.400705   0.035914 39.0014 < 2.2e-16 ***
Xadvanced city        0.118148   0.071828  1.6449 0.1062709
Xadvanced emerging   -0.143284   0.053269 -2.6898 0.0096917 **
Xarab oil-based       0.011441   0.056785  0.2015 0.8411486
Xcoordinated market  -0.175289   0.046365 -3.7806 0.0004182 ***
Xemerging            -0.133275   0.040949 -3.2547 0.0020393 **
Xeuropean peripheral -0.178720   0.045428 -3.9341 0.0002580 ***
Xhighly coordinated  -0.276081   0.095020 -2.9055 0.0054494 **
Xsocialist           -0.211703   0.071828 -2.9473 0.0048596 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    0.76517
Residual Sum of Squares: 0.38695
R-Squared:      0.4943
Adj. R-Squared: 0.41339
F-statistic: 6.10905 on 8 and 50 DF, p-value: 1.7649e-05
```

Table 23. World governance indicators vs. entropy model
Oneway (individual) effect Within Model

Call:
plm(formula = Y ~ X + factor(year) + factor(country), data = full_panel_data,
model = "within", index = c("country", "year"))

Unbalanced Panel: n = 196, T = 7-18, N = 3409

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-3.2694498	-0.1658852	-0.0047872	0.2102318	2.0042730

Coefficients:

	Estimate	Std. Error	t-value	Pr(>|t|)
XGovernment Effectiveness	-0.1224977	0.0484433	-2.5287	0.0114969 *
XControl of Corruption	0.1039579	0.0464007	2.2404	0.0251311 *
XPolitical Stability and …	0.0878272	0.0249842	3.5153	0.0004454 ***
XRule of Law	-0.1549039	0.0528882	-2.9289	0.0034258 **
XRegulatory Quality	0.1240954	0.0427152	2.9052	0.0036956 **
XVoice and Accountability	0.1716377	0.0418366	4.1026	4.188e-05 ***
factor(year)1998	0.0581267	0.0460222	1.2630	0.2066764
factor(year)2000	0.0046474	0.0460871	0.1008	0.9196850
factor(year)2002	0.1695499	0.0459649	3.6887	0.0002292 ***
factor(year)2003	0.2631962	0.0456116	5.7704	8.667e-09 ***
factor(year)2004	0.3481829	0.0454946	7.6533	2.578e-14 ***
factor(year)2005	0.4465587	0.0455014	9.8142	< 2.2e-16 ***
factor(year)2006	0.6154144	0.0453995	13.5555	< 2.2e-16 ***
factor(year)2007	0.6252912	0.0454158	13.7681	< 2.2e-16 ***
factor(year)2008	0.6203456	0.0454300	13.6550	< 2.2e-16 ***
factor(year)2009	0.7503412	0.0453032	16.5626	< 2.2e-16 ***
factor(year)2010	0.8005905	0.0453063	17.6706	< 2.2e-16 ***
factor(year)2011	0.8467331	0.0454075	18.6474	< 2.2e-16 ***
factor(year)2012	0.8917639	0.0453474	19.6652	< 2.2e-16 ***
factor(year)2013	0.9392799	0.0452399	20.7622	< 2.2e-16 ***
factor(year)2014	1.0153631	0.0454418	22.3443	< 2.2e-16 ***
factor(year)2015	1.0144551	0.0454390	22.3256	< 2.2e-16 ***
factor(year)2016	1.0664494	0.0454361	23.4714	< 2.2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares: 1015
Residual Sum of Squares: 600.38
R-Squared: 0.40851
Adj. R-Squared: 0.36808
F-statistic: 95.7879 on 23 and 3190 DF, p-value: < 2.22e-16

<u>Table 24. World governance indicators vs. types of economy</u>
- between estimator models

24.1 Control of corruption

```
Oneway (individual) effect Between Model

Call:
plm(formula = WGI ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 18-18, N = 1062
Observations used in estimation: 59

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.806313 -0.275292 -0.013627  0.232864  0.833237

Coefficients: (1 dropped because of singularities)
                     Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)          1.852376   0.170938  10.8365 1.005e-14 ***
Xadvanced city       0.093597   0.341876   0.2738    0.7854
Xadvanced emerging  -1.237214   0.253542  -4.8797 1.125e-05 ***
Xarab oil-based     -1.360720   0.270277  -5.0345 6.597e-06 ***
Xcoordinated market  0.172210   0.220680   0.7804    0.4389
Xemerging           -2.394691   0.194899 -12.2868 < 2.2e-16 ***
Xeuropean peripheral -1.306252  0.216221  -6.0413 1.888e-07 ***
Xhighly coordinated -0.483127   0.452259  -1.0683    0.2905
Xsocialist          -2.291957   0.341876  -6.7041 1.743e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    68.923
Residual Sum of Squares: 8.7659
R-Squared:      0.87282
Adj. R-Squared: 0.85247
F-statistic: 42.8915 on 8 and 50 DF, p-value: < 2.22e-16
```

24.2 Government effectiveness

Oneway (individual) effect Between Model

```
Call:
plm(formula = WGI ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 18-18, N = 1062
Observations used in estimation: 59

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.0349839 -0.2301303  0.0032029  0.1871065  1.2240810

Coefficients: (1 dropped because of singularities)
                   Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)          1.69667    0.16222  10.4588 3.477e-14 ***
Xadvanced city       0.20573    0.32445   0.6341 0.5289200
Xadvanced emerging  -0.87923    0.24062  -3.6541 0.0006187 ***
Xarab oil-based     -1.29966    0.25650  -5.0669 5.897e-06 ***
Xcoordinated market  0.17729    0.20943   0.8465 0.4012881
Xemerging           -1.90275    0.18496 -10.2871 6.149e-14 ***
Xeuropean peripheral -0.91988    0.20520  -4.4829 4.312e-05 ***
Xhighly coordinated  -0.27827    0.42921  -0.6483 0.5197352
Xsocialist          -2.39736    0.32445  -7.3890 1.480e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    50.484
Residual Sum of Squares: 7.895
R-Squared:      0.84361
Adj. R-Squared: 0.81859
F-statistic: 33.715 on 8 and 50 DF, p-value: < 2.22e-16
```

24.3 Political stability and absence of violence & terrorism

```
Oneway (individual) effect Between Model

Call:
plm(formula = WGI ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 18-18, N = 1062
Observations used in estimation: 59

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.328514 -0.362864  0.060153  0.310379  1.015672

Coefficients: (1 dropped because of singularities)
                    Estimate Std. Error t-value  Pr(>|t|)
(Intercept)          0.92900    0.22557  4.1185 0.0001429 ***
Xadvanced city       0.16697    0.45114  0.3701 0.7128690
Xadvanced emerging  -1.22413    0.33457 -3.6588 0.0006098 ***
Xarab oil-based     -0.46517    0.35666 -1.3042 0.1981240
Xcoordinated market  0.24029    0.29121  0.8252 0.4132031
Xemerging           -1.69301    0.25719 -6.5828 2.699e-08 ***
Xeuropean peripheral -0.32510    0.28533 -1.1394 0.2599718
Xhighly coordinated  0.11453    0.59680  0.1919 0.8485977
Xsocialist          -1.28241    0.45114 -2.8426 0.0064622 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    51.49
Residual Sum of Squares: 15.265
R-Squared:      0.70355
Adj. R-Squared: 0.65611
F-statistic: 14.8325 on 8 and 50 DF, p-value: 7.5409e-11
```

24.4 Regulatory quality

```
Oneway (individual) effect Between Model

Call:
plm(formula = WGI ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 18-18, N = 1062
Observations used in estimation: 59

Residuals:
      Min.     1st Qu.      Median     3rd Qu.        Max.
-0.65125134 -0.16291466  0.00057748  0.15925576  0.82315150

Coefficients: (1 dropped because of singularities)
                      Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)           1.688727   0.133729  12.6280 < 2.2e-16 ***
Xadvanced city        0.246501   0.267457   0.9216   0.36114
Xadvanced emerging   -0.861286   0.198352  -4.3422 6.879e-05 ***
Xarab oil-based      -1.360327   0.211444  -6.4335 4.618e-08 ***
Xcoordinated market  -0.077256   0.172643  -0.4475   0.65646
Xemerging            -1.923812   0.152474 -12.6173 < 2.2e-16 ***
Xeuropean peripheral -0.772846   0.169155  -4.5689 3.233e-05 ***
Xhighly coordinated  -0.632680   0.353813  -1.7882   0.07981 .
Xsocialist           -3.029012   0.267457 -11.3252 2.072e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    48.454
Residual Sum of Squares: 5.365
R-Squared:      0.88928
Adj. R-Squared: 0.87156
F-statistic: 50.1967 on 8 and 50 DF, p-value: < 2.22e-16
```

24.5 Rule of law

```
Oneway (individual) effect Between Model

Call:
plm(formula = WGI ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 18-18, N = 1062
Observations used in estimation: 59

Residuals:
     Min.    1st Qu.     Median    3rd Qu.       Max.
-0.8027046 -0.1564831  0.0088734  0.1582334  0.9291861

Coefficients: (1 dropped because of singularities)
                     Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)          1.733220   0.146314  11.8459 3.985e-16 ***
Xadvanced city      -0.203183   0.292628  -0.6943    0.4907
Xadvanced emerging  -1.055648   0.217019  -4.8643 1.186e-05 ***
Xarab oil-based     -1.300632   0.231343  -5.6221 8.411e-07 ***
Xcoordinated market  0.086087   0.188891   0.4557    0.6505
Xemerging           -2.202600   0.166824 -13.2031 < 2.2e-16 ***
Xeuropean peripheral -0.952422   0.185074  -5.1462 4.479e-06 ***
Xhighly coordinated -0.390290   0.387111  -1.0082    0.3182
Xsocialist          -2.893226   0.292628  -9.8870 2.353e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    59.172
Residual Sum of Squares: 6.4224
R-Squared:      0.89146
Adj. R-Squared: 0.8741
F-statistic: 51.3345 on 8 and 50 DF, p-value: < 2.22e-16
```

24.6 Voice & accountability

```
Oneway (individual) effect Between Model

Call:
plm(formula = WGI ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 18-18, N = 1062
Observations used in estimation: 59

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.141698 -0.147320  0.044174  0.200422  0.900020

Coefficients: (1 dropped because of singularities)
                    Estimate Std. Error t-value  Pr(>|t|)
(Intercept)          1.38785    0.18133  7.6537 5.727e-10 ***
Xadvanced city      -1.21367    0.36266 -3.3466  0.001559 **
Xadvanced emerging  -0.81581    0.26895 -3.0333  0.003830 **
Xarab oil-based     -2.37213    0.28671 -8.2737 6.294e-11 ***
Xcoordinated market  0.11137    0.23409  0.4757  0.636335
Xemerging           -1.83530    0.20675 -8.8770 7.551e-12 ***
Xeuropean peripheral -0.40653   0.22936 -1.7724  0.082420 .
Xhighly coordinated -0.36747    0.47975 -0.7660  0.447297
Xsocialist          -2.55463    0.36266 -7.0442 5.118e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    57.398
Residual Sum of Squares: 9.864
R-Squared:      0.82815
Adj. R-Squared: 0.80065
F-statistic: 30.1179 on 8 and 50 DF, p-value: < 2.22e-16
```

Entropy, GDP & specialization vs. types of economy - table 24
- between estimator models

24.7 Entropy

```
Oneway (individual) effect Between Model

Call:
plm(formula = entropy ~ X, data = panel.h4, model = "between")

Balanced Panel: n = 59, T = 21, N = 1239
Observations used in estimation: 59

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.773567 -0.085757  0.029195  0.173767  0.454862

Coefficients: (1 dropped because of singularities)
                      Estimate Std. Error t-value  Pr(>|t|)
(Intercept)            7.44993    0.10423 71.4786 < 2.2e-16 ***
Xadvanced city        -0.47443    0.20845 -2.2760 0.0271659 *
Xadvanced emerging    -0.30664    0.15459 -1.9836 0.0528097 .
Xarab oil-based       -0.60767    0.16480 -3.6874 0.0005584 ***
Xcoordinated market   -0.10670    0.13456 -0.7930 0.4315528
Xemerging             -0.66336    0.11884 -5.5822 9.688e-07 ***
Xeuropean peripheral  -0.33098    0.13184 -2.5105 0.0153348 *
Xhighly coordinated   -0.47331    0.27576 -1.7164 0.0922780 .
Xsocialist            -0.63005    0.20845 -3.0225 0.0039468 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6.7606
Residual Sum of Squares: 3.2589
R-Squared:      0.51795
Adj. R-Squared: 0.44083
F-statistic: 6.71557 on 8 and 50 DF, p-value: 6.0634e-06
```

24.8 GDP

Oneway (individual) effect Between Model

Call:
plm(formula = logGDP ~ X, data = panel.h4, model = "between")

Unbalanced Panel: n = 59, T = 19-21, N = 1236
Observations used in estimation: 59

Residuals:

```
    Min.  1st Qu.   Median  3rd Qu.     Max.
-2.10432 -0.71449 -0.24365  0.50432  2.68917
```

Coefficients: (1 dropped because of singularities)

```
                    Estimate Std. Error t-value  Pr(>|t|)
(Intercept)         27.49964    0.46102 59.6494 < 2.2e-16 ***
Xadvanced city      -1.58347    0.92204 -1.7173  0.092104 .
Xadvanced emerging  -1.08977    0.68380 -1.5937  0.117311
Xarab oil-based     -1.95417    0.72894 -2.6808  0.009921 **
Xcoordinated market -0.70278    0.59518 -1.1808  0.243272
Xemerging           -1.35257    0.52565 -2.5732  0.013091 *
Xeuropean peripheral -1.14353   0.58315 -1.9610  0.055466 .
Xhighly coordinated  1.70394    1.21975  1.3970  0.168595
Xsocialist          -2.28332    0.92204 -2.4764  0.016701 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Total Sum of Squares: 87.775
Residual Sum of Squares: 63.762
R-Squared: 0.27358
Adj. R-Squared: 0.15735
F-statistic: 2.35378 on 8 and 50 DF, p-value: 0.031175

24.9 Specialization - Hoover coefficient

Oneway (individual) effect Between Model

```
Call:
plm(formula = hoover.specialization ~ X, data = panel.h4, model = "between")

Balanced Panel: n = 59, T = 21, N = 1239
Observations used in estimation: 59

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.147073 -0.044098 -0.011234  0.034362  0.185814

Coefficients: (1 dropped because of singularities)
                      Estimate Std. Error t-value  Pr(>|t|)
(Intercept)          0.1684926  0.0295012  5.7114 6.127e-07 ***
Xadvanced city        0.0879857  0.0590024  1.4912  0.142185
Xadvanced emerging    0.0731625  0.0437573  1.6720  0.100771
Xarab oil-based       0.1553500  0.0466455  3.3304  0.001635 **
Xcoordinated market  -0.0200838  0.0380859 -0.5273  0.600297
Xemerging             0.1614246  0.0336365  4.7991 1.482e-05 ***
Xeuropean peripheral  0.0248521  0.0373164  0.6660  0.508483
Xhighly coordinated   0.0095682  0.0780528  0.1226  0.902926
Xsocialist            0.1662783  0.0590024  2.8182  0.006901 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    0.59547
Residual Sum of Squares: 0.2611
R-Squared:      0.56153
Adj. R-Squared: 0.49137
F-statistic: 8.00401 on 8 and 50 DF, p-value: 7.106e-07
```

F. R Scripts

For R scripts see separately attached file.