LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Growth,

Innovation and Spatial Dynamics

The gravity of tacit knowledge flows in the presence of Information Communication Technologies

Kristofer Rolf Söderström

kristofer_rolf.soderstrom.064@student.lu.se

Abstract: The role of geographical distance on the flow of knowledge has been debated in empirical work over the past decades. The magnitude of this relationship is usually attributed to the type of knowledge in question; in which tacit knowledge—the focus of the thesis—is usually associated with a negative relationship. The purpose of this thesis is to evaluate said role in the context of tacit knowledge flows, and in the presence of the diffusion of Information Communication Technologies (ICTs). Using data for international patent cooperation as proxy for tacit knowledge flows, a gravity model is implemented for an analysis of 48 countries and period 1993-2012. Although several attempts were implemented with numerous approaches, the presence of a negative relationship between geographical distance and tacit knowledge flows could not be contested within the context of ICT diffusion. This suggests evidence for the important role of geographical distance in the environment of face-to-face communication. The findings support current theoretical argumentations within the literature.

Key words: Gravity, tacit knowledge, knowledge flows, ICT, diffusion, PPML, patents, cooperation, distance, attractors, unit root, cointegration.

EKHM52
Master thesis, second year (15 credits ECTS)
June 2016
Supervisor: Moodysson, Jerker
Examiner: Josef Taalbi
Word Count: 15 344

Table of Contents

List of Tables, Figures & Appendix

Tables

Figures

Appendix

1 Introduction

1.1 Research Problem

The determinants of knowledge flows and the role of geographical distance are topics of interest within the literature regarding the knowledge economy. Focusing on tacit knowledge, several theoretical arguments can be made to justify the importance of geographical proximity or face-to-face communication between individuals. However, empirical work on this feature is still ongoing with different arguments for either side of the debate. Thus, an attempt is made to provide some of this needed empirical backing to tacit knowledge flows.

Information Communication Technologies can be seen as one of the many foundations for the knowledge economy. Facilitating communication between economic agents has arguably enabled much needed communication in the realm of productivity and economic growth. The frontiers of this influence are however, much debated in their current state. The success of the communication seems to hinge on the type of content that is desired to transmit. Thus, a central point of the thesis is to study the relationship of distance and knowledge flows, tacit in nature, in the presence of ICTs.

However, geographical proximity has found robustness in the relationship with most types of interactions between entities. This negative relationship has been found not only in trade theory or the flow of knowledge, but in many areas of research. Thus, of particular interest is to understand the role of Information Communication Technologies in this context and to see if the increase of "long distance" communication has any kind of influence on the role geographical distance.

The thesis will use proxies for tacit knowledge flows as a dependent variable; and proxies for ICT access and diffusion, geographical distances and for cultural ties as explanatory variables. Then, an analysis of these factors between country pairs and across a period will be performed in order to identify possible influences and associations with the dependent variable. The estimation will be performed with a gravity type equation between country pairs, in which bilateral flows are measured in terms of attractors and measurements of distance.

1.2 Aim and scope

If one accounts for the role of ICT diffusion in the analysis, it may be possible to provide some insight into to whether the theoretical notion about the ineffectiveness of ICT in communicating tacit knowledge across large distances is correct. Or rather, increasing the number of possible methods of communication by which individuals between countries can connect may incentivize an exchange of tacit knowledge, regardless of distance. The motivation behind the inclusion of the indicators is two-fold. Firstly, it is interesting to provide evidence to whether the presence of communication technologies between the countries explains some of the flows of tacit knowledge. Secondly, and the main question in the thesis, is whether this presence is able to reduce the effect of geographical distance on the flow of face-to-face knowledge.

The research area relates to the determinants of knowledge flows, specifically tacit knowledge. A proxy for tacit knowledge flows is identified for the study, and will be measured as international cooperation between countries at the inventor level. It is expected that several factors may be responsible for a successful cooperation between creators. Thus, it is the interest of the thesis to understand the relationship and significance, both in statistical and economic terms, of these factors on the flow of tacit knowledge between individuals.

1.3 Outline of the thesis

The outline of the thesis follows the following structure: Section 2 presents the theory used for the purposes of the thesis. First by a review of the literature in a deductive composition, starting from a general overview of the role of knowledge in the economy, up to the recent attempts to employ gravity type models in a similar context to the thesis, followed by the theoretical structure of the gravity model itself and some of its background. Section 3 presents the data used in the analysis. It provides a description, a priori expectations and limitations, as well as the description of the necessary transformations for the data. The data is divided by the different sections of the gravity equation in order to provide some clarity and compartmentalization of the concepts. Section 4 presents the methods in which the analysis of the thesis is implemented. First, the model specifications for the econometric estimations are presented, which will follow the gravity framework. Then, a short description of the employed estimation method and its motivation is provided, since it differs from traditional OLS estimations. Section 5 provides the empirical analysis for the thesis, first by providing descriptive statistics and additional aids to provide a general acquaintance of the dependent and explanatory variables. Then, the output from the econometric estimations is presented, followed by a discussion of the results themselves. Finally, Section 6 concludes the thesis and includes some recommendations for future research.

2 Theory

2.1 Previous Research

Studying the role of technology and knowledge in the economy has had a relatively recent surge in interest by economists. According to Samuelson (2004), "throughout most of the history of economics, the question of who knows what was considered unimportant…It is now clear that much of economic importance depends upon what people know"(pp. 368). Criticizing the limits of traditional neoclassical models and appreciating the need to include technological change into models for economic growth is usually first attributed to Solow (1956). Additional studies followed suit, which stress this issue and include, but are not limited to, Romer (1986) and Lucas (1988) who argue for the importance of knowledge in the economic process, as investment in knowledge and human capital can aid to the creation of growth, endogenously, through the transfer of knowledge between firms and industries.

The field has been in development for some time, which has put some of the debate to rest; it is now clear that there are different types of knowledge with different characteristics and methods of creation and transfer or diffusion. Malecki (2010) states that there is an importance in the mechanisms of the transmission of knowledge and how it is composed, between which actors it is transferred, at which limits of distances does this transmission hold and in which basis and codebooks, as well as the real impact of research facilities and local universities on innovative activity. Social relationships are highlighted as they are of relative importance for the author; since it may aid in explaining flows of knowledge; enduring over space, time and organizational boundaries. Indeed, the inner workings of the transmission of knowledge have many different dimensions that make their study in need of nuance.

New and improved methods of measuring knowledge are in need in order to provide a better understanding of this complicated phenomenon. One aspect of knowledge is studied by Moodysson (2008), which argues for a rethinking of the approach to interactive knowledge creation. In an analysis of the Medicon Valley life science region in Sweden with the insights from the study of clusters and innovation systems, the author finds that local buzz seems to be absent in the mechanisms of knowledge flows within the area. Rather, most interactive knowledge seems to be spontaneous and unregulated, safely embedded in a professional community of knowledge, globally accessible only by those who qualify.

The role of Information Communication Technologies in the economy has long been studied in the past decade. ICT is seen as a tool that contributes to rapid technological progress and productivity growth (Seki, 2008). The ability to use these technologies to solve problems, working in teams, supervise and to undertake continuous learning is seen as crucial elements for the labour market. Furthermore, the increased growth in knowledge-based industries takes place at a moment in time when investment for ICT technologies and the growth, and use, of the internet have increased (OECD, 2001). However, the role of ICT in knowledge flows is not as straightforward, since not all types of knowledge behave the same, and will not be communicated efficiently in similar methods.

Thus, it is important to clarify that for the purposes of the thesis, the interest is shifted towards tacit knowledge, which contrary to codified knowledge, usually requires face-to-face interaction between actors. This type of knowledge relates to the "know-how" aspect, and has been extensively studied in the literature. Tacit knowledge is usually regarded as difficult to codify, and is not possible to reduce to numbers, graphs, formulae, and other types of codification with a need of physical communication. Innovativeness and competitiveness of firms, regions and nations are in need of interaction, networking, co-operation, social capital and spatial proximity, which all constitute elements for collective the learning process.

Face-to-face interaction is taken literally as the interaction between two people—or more—which allows for visual and physical contact. Thus, interactions that are done electronically are by definition, not a component of tacit knowledge (Asheim et al. 2005). In other words, it can be argued that it traditionally excludes Information Communication Technologies as a viable communication method in this context.

Asheim et al (2005) point out that from the literature, it can be seen that tacit knowledge has been receiving increasing importance in the learning economy due to its prominence in innovation via interacting learning. The authors call attention to the lack of empirically sustained work on this topic. Thus, relevant research has to be increased in order to provide empirical support to the theoretical aspect of the mechanisms for the transfer of tacit knowledge.

Studies on the flow of tacit knowledge are usually done with patent citations as proxy. A study by Peri (2005) reveals that knowledge is highly localized within the sample. In the study of 1.5 million patents and 4.5 million patent citations across 147 subnational regions for the period 1975-1996, the author finds that only 20% of the average knowledge is learned outside the regional level and that only 9% is learned outside the country of origin. The author also finds that knowledge that comes from technologically leading regions reaches farther distances than the average and that knowledge from the computer sector behaves in a similar fashion. The estimation is made through a gravity-like equation with a negative-binomial specification. Knowledge flows, with patent citations as a proxy, seem to be highly localized. This seems to be comparable to some of the other studies in this section, which seem to conclude that distance does seem to be an important factor when analysing patent citations. The authors also compare their work with Jaffe et al. (1993), which uses a smaller sample for the case of the United States; and with Maurseth and Verspaguen (2002), which use citations between patents in 112 European regions. Both studies find similar effects on the effect of geographical distance on patent citations, although at different levels with the latter (Peri, 2005).

This provides an interesting perspective on the effect of distance on patent citations. There seems to be some kind of robustness in its effect across some of the patent citation literature for knowledge flows, although the total extent of the literature cannot be reviewed in this section. This provides some precedent in the possible effects of distance in the thesis. Although using a different measurement for tacit knowledge flows, the effect of distance is expected to be similar.

Patent citations are not the only measurement in the literature regarding the flow of tacit knowledge. Another one is international cooperation in patents and a gravity model is applied in this context by Picci (2010), in which an analysis is implemented on the determinants of internationalization of patent applications in Europe between 1990 and 2005. The author defines the term in three ways depending if the registration belongs to a domestic inventor with a foreign applicant; a domestic applicant with a foreign inventor; or a co-ownership between inventors in domestic and foreign countries. Thus, the author measures collaboration between inventor and applicant, applicant and inventor, and inventor and inventor, in the country of origin and the partner country, respectively. The author finds that internationalization has been increasing over time, although it is rather low in its totality. Using a gravity model to account for the bilateral collaboration between countries, the author measures the effect of geographical distance, common language, common border and other cultural characteristics.

The results from Picci (2010) indicate that geography and culture play an important role in the determinants of internationalization of patents over the period of 1990-2005 and in different subsets of countries. The author does note that the period is rather short and limits the selection of countries to OECD members and the European Union. It should be noted however, that the by Picci (2010) uses the natural logarithmic transformation for the dependent variable. This is a problem because the dependent variable is a count measurement in which the number of observations ranges from zero to infinity. The proportion of zeroes is also naturally high and not random, i.e. a zero count between two countries in the data is a result in itself and should not be eliminated. Eliminating zero values, which is the case when taking natural logarithm, will result in a bias in the estimation and obviously result in a much lower number of observations. For this reason, the results from this article should be taken with consideration.

Montobbio & Sterzi (2013) have published another example that measures international patent collaborations. The focus on this article is on the determinants of collaboration between patent inventors in emerging and advanced countries. Using the patent database from the US Patent and Trademark Office (USPTO), the authors collect data for eleven emerging economies and seven advanced countries for the period 1990-2204 and estimate the impact of several distance measurements, economic and institutional variables on collaboration using the Poisson Pseudo-Maximum Likelihood (PPML) Estimator. The authors find several results depending on the type of collaboration and found different effects between emerging and advanced countries. With respect to the interests of the thesis, the main results are the following: Geographical distance is not important by itself, but find significant effects through trade channels and cultural similarities and find stronger results for time zone differences (latitude distance). Montobbio & Sterzi (2013) also measure distance in terms of technological proximity with a method attributed to Jaffe (1988), in which the un-centered correlation of each country pair's vector of patents across different technological classes and ranges from 0 to 1 for all pairs. The authors find this variable to be a determining factor, which reduces the effect of geographical distance in collaboration when also controlling for common language. The authors, in contrast with Picci (2010) the authors of this study do take into consideration the nature of the dependent variable and appropriately use the PPML estimator with the dependent variable in levels.

Montobbio & Sterzi (2013) have discussed qualifying patent cooperation as a proxy for knowledge flows. The authors propose that international co-operation between inventors may be subject to different types of activities, and measure different types of knowledge flows. The authors cite an example of a study on the firm ABB, in which 53 inventors are interviewed on the nature of the collaboration within the firm and find that around 60% of inventions are an actual result of international collaboration; while half are the result of international R&D activities and only one third are joint R&D projects. Almost all of the residual 40% are due to inventor movements. Regardless, the authors believe that this is an appropriate proxy for knowledge flows, specifically tacit knowledge.

In a study by Ejermo and Karlsson (2006) of interregional inventor networks in Sweden, aims to explain the structure and strength of inventor and co-inventor networks, measured as affinity, when co-authoring patents across Sweden. The authors find that in general, travel distance has a great influence in affinity. However, affinities extend more to regions with high patenting and R&D levels, and those to University R&D. Furthermore, regions with high R&D tend to look inwards, in a relative sense.

In a paper by Steinmueller (2000), the author argues for the potential of Information Communication Technologies on the codification of knowledge and its possible improvement over time. The most interesting argument in terms of the thesis is the one for the World Wide Web (WWW). The author mentions that since its development, the variety of formats in which information can be exchanged has increased greatly. It is interesting to note that in the time be after the release of the article by Steinmueller (2000), said variety must have increased in many more times. Although the author seems to deal with codified knowledge, it may be possible to assume that these developments could have some consequence in tacit knowledge as well.

Considering the previous research as a stepping-stone for the thesis, it may be possible to construct a research question that would aim to provide more information on the relationship between tacit knowledge flows, geographical distance and Information Communication Technology. Specifically, to understand the possible effects of Information Communication Technologies on the dynamic between tacit knowledge flows and distance. Thus, the ***research question*** is as follows:

"Does the diffusion of Information Communication Technologies negate the influence of geographical distance on tacit knowledge flows at the inventor level?"

Several hypotheses are constructed in order to provide evidence for the research question. The hypotheses are constructed with the same rejection structure and are related to the following points of interest, followed by a short description:

1. On the relationship between geographical distance and tacit knowledge flows

H0: An increase in geographical distance does not have a negative association with tacit knowledge flows.

H1: An increase in geographical distance has a negative association with tacit knowledge flows.

A first step in the method is to construct a simple, non-augmented, gravity model to understand the baseline characteristics of geographical distance. A model that is related to the traditional gravity equation in classical mechanics, it only includes the attractors and the measurement of distance measured in kilometres between country pairs[1].

2. On the effect of the diffusion of Information Communication Technology on the relationship between geographical distance and tacit knowledge flows.

H0: The diffusion of ICT does not have a negative effect on the association between geographical distance and tacit knowledge flows.

H1: The diffusion of ICT has a negative effect on the association between geographical distance and tacit knowledge flows.

The second step is to include the possible effect of Information Communication Technologies (ICT) in the augmented gravity model. The model will include the parameters from the non-augmented model, with different measurements for ICT for robustness purposes in the model. As with the non-augmented model, control variables are included to reduce the possibility of omitted variable bias in the model.

The intuition behind the research question and the hypotheses lies on understanding the relationship between geographical components and tacit knowledge flows in a first instance, following the thread from the previous research. This will allow the thesis to align itself with the results from the previous literature that has found, for the most part, a negative relationship. After this initial step, the diffusion of ICT is included in the model in order to understand how this may have an influence on the dependent variable. If correct in the method, model specification and data, perhaps some insight may be found on the role of ICT in tacit knowledge flows.

However, the effect of the internet may very well be the main interest. As tacit knowledge flows usually requires face-to-face interaction, perhaps it is more likely that the flexibility of the internet will make it more likely that this characteristic of tacit knowledge depend less on geographical proximity. Thus, an additional hypothesis is included in the thesis to account for this possibility.

3. On the effect of the diffusion of Internet on the relationship between geographical distance and tacit knowledge flows.

[1] An expansion of this model is also included in which other variables are added to account for additional measurements of distance, details will follow suit.

H0: The diffusion of Internet does not have a negative effect on the association between geographical distance and with tacit knowledge flows.

H1: The diffusion of Internet has a negative effect on the association between geographical distance and with tacit knowledge flows.

The study contains an initial sample of 48 countries across the period 1993-2012. Since the interest is to measure the role of distance and ICT technologies in the flow of tacit knowledge in a general aspect, it seems reasonable to include as many countries as possible in the analysis. This would attempt to overcome any biases regarding the selection of developed versus developing countries, or members of common economic agreement, large countries or small and remote islands.

However, only increasing the number of countries does not automatically allow the results to be inferred in the sense of a general equilibrium. Still, some precautions have to be taken into the interpretation of the results. There may be issues regarding the selection of data, model specification, method of estimation, biases in the sample, amongst many others. Nevertheless, most of these issues will be addressed, as much as possible, in order to decrease the likelihood of committing statistical errors.

2.2 Theoretical Approach

The aim of the thesis, as mentioned earlier, is to measure tacit knowledge spillovers between countries at the inventor level. Using international cooperation in patenting as proxy for tacit knowledge flows, measured by means of co-invention, brings the particular challenge of measuring bilateral exchange of ideas, while including factors that may influence or determine this cooperation. Thus, it seems reasonable that the application of a gravity model that measures this effect is the most appropriate to use.

Although traditionally employed for the analysis of trade flows between countries, it has had some success in other fields. From its conception in classical mechanics, to applications in migration flows, to economics, it is an interesting tool to measure bilateral flows between countries. It also allows for the inclusion of a time dimension, making it even more powerful. Fortunately, there is precedent in the literature to apply this method in the context of patent cooperation.

Gravity models have had relative success in empirical studies, it is "…considered one of the most successful empirical frameworks in international economics" (Herrera, 2013; pp. 1101). It aims to measure the "pull" between two masses in terms of attractors and distance. The model has its basis in Newton's law of gravitational force (GF) between two objects which can be expressed in the following equation form:

$$GF_{ij} = \frac{M_i M_j}{D^2{}_{ij}} \; i \neq j \quad \textbf{(1)}$$

Where M equals to mass and D equals distance between objects $i \neq j$. In equation 1, gravitational force is directly proportional to the masses of the objects and indirectly proportional to the distance squared between them. The first theoretical introduction to economics dates back to Jan Tinbergen in 1962 in which the model was considered as potentially applicable to the field of economics because of its useful analogy with "fortunate empirical validity" (Reinert et al., 2010; Pp.568) and was the first time to be applied to trade flows (Anderson, 2010).

Over the next few decades, links with elements in trade theory had been implemented and further revisions to the model improved its explanatory power. The first attempt to provide a theoretical base of the model in economics was done by Anderson (1979), while empirical studies where proved to work in the past, now it is known that most trade models need a gravitational specification in order to function correctly. Thus, it is a usual point of departure for the contemporary specification of the model (Reinert et al., 2010; Anderson, 2010; WTO, 2012).

In its most basic and intuitive form, the gravity model for international trade can be adapted from Shepherd (2013) and Reinert et al. (2013) as the following equation:

$$\ln(X_{ij}) = \propto + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) + \beta_3 \ln(D_{ij}) + \varepsilon_{ij} \quad \textbf{(2)}$$

Where X_{ij} refers to exports from country i to country j; GDP refers to the gross domestic product for each country; D_{ij} represents some indicator of bilateral distance variables between each country pair, such as geographical distance, cultural ties, tariffs, or others; and ε_{ij} is a random error term. This specification is usually expanded in more recent studies to include a time dimension for a dynamic analysis of gravity models. Traditionally, logs of the dependent and independent variables are taken in order to measure the elasticities of the coefficients. However, this is not always advisable due to the possible loss of information from the transformation of data when the indicators are count variables. For this reason, several modifications will be made to the original gravity model for bilateral trade. These modifications will take some inspiration from the previous research, but an attempt will also be made to include some degree of originality and uniqueness to the thesis.

According to Head & Mayer (2013), the gravity equation can be applied to a wide range of bilateral flows and interactions, with key ingredients being a measurement of "mass" effects and bilateral and multilateral resistance terms. As soon as the equation is specified, the gravity models can usually be estimated using the usual techniques appropriate for measuring trade flows. Thus, it seems that the estimation of a gravity equation is warranted for the purposes of the thesis.

3 Data

3.1 Selected variables

The original database includes data that had been recollected for an initial count of 92 countries across the world and for the period 1980-2012. The countries include developed economies, emerging countries, OECD members, members of the European Union, amongst others. An attempt was made to include, at least initially, as much countries as possible in order to increase the scope of the thesis to provide a more general result in terms of tacit knowledge flows in an international perspective. However, it was necessary to reduce the original database in order to perform the necessary statistical testing to the variables. The reductions are specifically linked to statistical testing for unit root and cointegration between the variables and will be addressed accordingly[2]. These led to a final sample of 48 countries and period 1993-2012.

The dependent variable of interest is "International co-operation in research" ($coinv_{ijt}$) from the OECD (2016) indicators of international co-operation database. The variable measures international co-operation in patenting between inventors from different countries of residence, thus it is a measure of **co-invention** (inventor-inventor), rather than co-ownership in an international sense (inventor-foreign ownership or domestic ownership-inventor). The indicator can measure either international collaboration by researchers within a multinational corporation with research facilities in several countries, or by cooperation between researchers among several firms or institutions. Which reflects the purpose of the selection of this indicator as the dependent variable, as it "also reflects international flows of knowledge" (OECD, 2009; pp-128). The unit of measurement is number of patent applications with collaboration registered at the Patent Cooperation Treaty (PCT), in which applicants can simultaneously seek protection for inventions in 148 countries around the world, from 1978 onwards. This is more appropriate for cross-country comparisons since the measurement is relatively free of the "home advantage" bias, in which domestic applicants are more likely to fill patents in their home countries or regions (OECD, 2009; WIPO, 2016).

[2] See appendix A.1 for a full list of countries and appendix 3 for the final sample selection

The type of patenting collaboration is between an inventor in country i with at least one other inventor in a foreign country j from the total number of patents that were invented domestically. Thus, the share of international co-inventions from the total of domestic inventions for country i is then:

$$\frac{\sum_{j=1}^{N} P_{i,j}}{P_i} \quad \textbf{(3)}$$

Where P is a measure of patent applications by Priority Date, defined as the first filling worldwide and the closest date to invention and which according to the OECD (2009), is the best reference to measure inventive activity and performance from a technological, or economic point of view. Patent applications have a usual delay of 12 months after the priority and the grant is typically awarded after 3 to 5 years, which may include significant time lags in the data.

Since the interest is to measure the cooperation between two inventors (co-invention) in different countries, as proxy for tacit knowledge flows, and not its actual commercial success or impact in economic growth, or whether or not the application is granted, it seems reasonable to follow this advice.

3.2 Attractors

The attractors, i.e. the "mass" equivalent, in the gravity model include measurements for patent applications, gross domestic product and population. The idea behind the attractors is to find linkages between the country pairs that would make it likely for two inventors to cooperate with each other. In this framework, the size of the innovative activity and the size of the economy are expected to have an effect on the probability of the cooperation between two countries, in a similar outline to Montobbio & Sterzi (2013), which include patents and the labour force, and to Picci (2010), which limits the attractors to patent applications. The variables are measured in origin and partner country, and can be found in the OECD Patent Database (2016) and the World Development Indicators (2016) from the World Bank:

1. Patent Applications (pat_{it} & pat_{jt}): Number of patent applications to the Patent Cooperation Treaty (PCT), by priority year. The main attractor for international collaboration in patents, it is expected that inventors with residence in highly inventive countries will be more likely to cooperate with each other than with countries with less inventive activity. Evidently, it is necessary to measure this indicator in the same terms as the dependent variable in order for the results to be comparable.

2. Gross Domestic Product ($gdptot_{it}$ & $gdptot_{jt}$): Constant 2005 U.S. Dollars, constructed variable from GDP per Capita data and Population data. Similar to inventive activity, it is expected that the size of economic activity in a country is another likely attractor between inventors in different countries. Larger economies are more likely to be engaged with other economies of similar size through different types of collaboration, which would have a positive relationship with the dependent variable.
3. Population ($poptot_{it}$ & $poptot_{jt}$): Number of total inhabitants. A final attractor between inventors may be country size. However, the a priori expectations are not as strong as the former two as the size per se of a country may not necessarily influence international cooperation directly. The variable is still included to serve as a control for population size.

3.3 Information Communication Technologies (ICT)

The ICT indicators are freely available from the United Nations (2016) databank, which in turn report the figures by the World Telecommunication/ICT Indicators Database. The aim of the indicators is to proxy for the diffusion of Information Communication Technologies (ICT) in the countries of interest. The indicators will face some transformations in an attempt to find robust measurements of "distance" in the diffusion of ICT technologies between countries. For now, the variables will be explained in their base form:

1. Percentage of individuals using the internet (int_{it} & int_{jt}): Percentage of the number of reported internet users by country, relative to population size, in national surveys.
2. Fixed-telephone subscriptions per 100 inhabitants (tel_{it} & tel_{jt}): Percentage of the number of subscriptions to telephone networks with registered activity in the past three months, relative to population size.
3. Mobile-cellular telephone subscriptions per 100 inhabitants (mob_{it} & mob_{jt}): Percentage of the number of subscriptions to a public mobile telephone service with access to Public Switched Telephone Network using cellular technology with active SIM cards and excluding mobile data, relative to population size.

3.3.1 Proxy for the diffusion of Information Communication Technologies

Due to the purposes of this thesis and the theoretical approach in a gravity context, it may be necessary to perform a transformation to the indicators for ICT. The original data for ICT is in terms of percentage, users and active subscriptions, and it is included for both country of origin and partner country. However, there may be issues with this initial approach. Having data for each of the countries is traditionally reserved for the "attractors". The intent is not to only measure the indicators this term, but also in terms of the difference in diffusion patterns of ICT technologies between countries, in an attempt to include another measure of distance over the time period. Thus, the following transformation is executed in order to measure this interaction between country pairs, and is treated as proxy for difference in diffusion patterns.

The new indicators are calculated by subtracting the percentage values of the origin countries with the percentage values of the partner country, the difference between countries is then calculated in absolute terms in order to measure absolute distance and avoid negative values, since the direction of distance is not of interesting in this context. This initial transformation will be done for the data between country pairs and for three indicators. It is anticipated that this would measure "distance" between countries in terms of ICT diffusion patterns. We can illustrate and example; in 1997, internet users in Sweden accounted for 23,73% of its population while in Denmark, this amounted to 11,36%. Thus, the static difference in diffusion was of 12,35% for that year.

The ICT calculations are included as an additional model specification in order to provide a different perspective in the measurement of the role of ICT in the model specifications. The idea is inspired by Peri (2005), in which the author calculates the difference in technological advancement in terms of R&D spending per worker between the origin and partner country.

3.4 Geographical distance and cultural similarities

Several time invariant indicators are included to account for geographical, language and cultural characteristics between country pairs in the models. This aims to measure different types of distance and cultural differences between countries. All of the indicators can be found in the research centre CEPII website, which recollect indicators in country pairs for use in gravity type models (Mayer & Zignago 2011; Head & Mayer 2013).

1. Distance ($dist_{ij}$): Simple distance between the most populated cities in each country pair, calculated with the great circle formula, measured in kilometres. Instead of measuring capital cities, it is expected that cities with more inhabitants will be more likely to innovate more, relative to cities with less inhabitants in the same country and thus increase the chances of international cooperation. There were several distance measures available but this was thought to be the most appropriate. Nevertheless, a simple correlation analysis shows that there is a high correlation with

the other distance measures, including distance between capital cities and weighted distance measures in relation with population.

2. Contiguity ($contig_{ij}$): Dummy variable in which 1 indicates that a country pair shares a border and 0 otherwise. It is expected that sharing a common border will increase the probability of cooperation between the inhabitants of the country pairs to a similar degree than the simple distance between countries, but with some nuance. This variable will identify the number of neighbours for each country. Remote areas are expected to have more difficulties in cooperation.
3. Time zone ($tdiff_{ij}$) : Time zone difference between country pairs, measured in hours. Individuals may be as likely to cooperate with each other to work "around the clock" in an invention, or as to only cooperate if they share somewhat similar working hours during the day. This variable complements simple distance, which may not necessarily be sufficient to measure the effect of geographical distance.
4. Common language ($comlang\_off_{ij}$): Dummy variable in which 1 indicates that a country pair shares at least one official language and 0 otherwise. Sharing a common language is expected to increase the probability of cooperation between two inventors. The rise of English as a second language however, may make it likely that this indicator becomes of less importance over time.
5. Second common language ($comlang\_ethno_{ij}$): Dummy variable in which 1 indicates that a country pair shares one language that at least 9% of the population can speak and 0 otherwise. This indicator is hoped to act as a proxy for common second language, although it has obvious drawbacks. It may not necessarily reflect a common language between inventors and may or may not have an effect in international collaboration.
6. Colonial ties ($colony_{ij}$): Dummy variable in which 1 indicates that a country pair has ever been in a colonial relationship and 0 otherwise. Having colonial ties with a country may act as a proxy for some type of cultural similarity between countries. However, it may be overstating the importance, but more like an additional control that may explain part of the cooperation between inventors.

An additional control variable was constructed in order to control whenever the country pairs where the same country, i.e. Sweden and Sweden, which is expected to bias results. Although the dependent variable already accounts for this issue, it was necessary to create for all other analyses.

7. Same country ($same_{ij}$): Dummy variable in which 1 indicates that a country pair contains the same country as partner and 0 otherwise.

It is noteworthy to mention that there may be certain limitations when explaining decision making at the firm or individual level (co-invention) with the current selection of macroeconomic variables. It may be possible that another effect is being captured by the selected variables; such as other types of cross-country economic flows, which would not necessarily reflect the desired outcome. Not including a set of firm level or individual level variables proves to be a possible limitation in the thesis and should be kept in mind when evaluating the results, since potentially important information may be left out of the model.

4 Methods

4.1 Model specification

As mentioned earlier, a gravity model is used for the analysis in this thesis. It is suspected that a bilateral gravity model will be superior for studying the possible determinants of the internationalization of patent cooperation and for the purposes of addressing the research question. Similar to Montobbio & Sterzi (2013), results from a non-augmented model, which only includes mass and distance, are calculated initially to evaluate the model significance of gravity-like models for patent co-invention between countries. Subsequently, additional augmented models will include the additional indicators that will account for mass, distance, diffusion of Information Communication Technologies (ICT) and further transformations, and a set of additional language and cultural indicators.

The first model—**equation 4—**is an adaptation from equations 1 and 2, in which the non-augmented gravity model measures the relationship between the attractors and simple distance with the dependent variable, it can be specified as:

$$coinv_{ijt} = \beta_0 + \beta_1 \ln(pat_{it}) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8 \ln(dist_{ij}) \quad \textbf{(4)}$$

Where *patent applications*, *GDP* and *population* refer to the attractors in the model with data for both the origin (i) and the partner (j) country across time (t); and where *distance* is the time invariant geographical distance indicator between country pairs (ij). As mentioned earlier, the purpose of the non-augmented model is to measure the sole relationship of the attractors and distance with the patent co-invention indicator in the model.

The second model specification is defined in **equation 5**, which will include the rest of the indicators. The indicators for ICT diffusion however, will be added in their base form and in the form of attractors for origin (i) and the partner (j) country for time (t). This aims to test for robustness of the role of ICT and the distance measurements in the model specification before any transformation. It may be the case that the transformation does not reflect the desired reality of diffusion, so it should be handled with care. Thus, the second model can be specified as:

$$coinv_{ijt} = \beta_0 + \beta_1 \ln(pat_{it}) + \beta_2 \ln(pat_t) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8(int_{it}) + \beta_9(int_{jt}) + \beta_{10}(tel_{it}) + \beta_{11}(tel_{jt}) + \beta_{12}(mob_{it}) + \beta_{13}(mob_{jt}) + \beta_{14} \ln(dist_{ij}) + \beta_{15}(contig_{ij}) + \beta_{16}(tdiff_{ij}) + \beta_{17}(comlang\_off_{ij}) + \beta_{18}(comlang\_ethno_{ij}) + \beta_{19}(colony_{ij}) \quad \textbf{(5)}$$

Where *internet use, telephone access* and *mobile access* are added in their base levels, since the data is a proxy for diffusion, it is added to the attractors of the non-augmented model (equation 4) and also include data for both the origin (i) and the partner (j) country for time (t). D*istance*, *contiguity* and *time zone difference* are part of the time invariant geographical distance indicators between country pairs (ij); and common official language, *common language share by 9% of the population* and *colony* are part of the time invariant cultural indicators between country pairs (ij).

The main model—referred to as **equation 6**[3] in the results—replaces the indicators for ICT diffusion in levels with the transformation for the ICT indicators to measure difference in diffusion patterns. Where, in comparison with equation 5, the variables for the difference of *internet, telephone* and *mobile,* are included in the model and are measured between country pairs and over the period (ijt), while the rest of the model remains the same.

As mentioned in the section "Research Question", perhaps it may be more likely the pattern of internet diffusion between country pairs is the one responsible for the potential reduction of the effect of distance in tacit knowledge flows. For this reason, and additional model **(equation 7)** will also be calculated, in which the role of internet is isolated. In this model, the relationship will be measured only in terms of the difference in the pattern of diffusion for internet use, omitting the variables for telephone and mobile from the previous model. Another motivation lies with the results of the correlation analysis. The high degree of collinearity between the ICT indicators may provide an undesirable effect on the results from previous

[3]See Appendix 2 for a list of the model specifications

model specifications. Furthermore, since the effect of internet may be the only one reflecting better communication quality over time, it is of special interest in the thesis.

However, it may be possible that omitting the telephone and mobile phone variables of ICT would result in neglecting important information in the model. Thus, a final model **(equation 8)** specification is in order, in which a final attempt is made to measure the role of ICT as a whole, where the average difference in diffusion patterns of ICT ($avgICT_{ijt}$) is calculated as the average of the three indicators in equation five (5). This aims to capture the difference in diffusion patterns of the whole of ICT indicators, regardless of origin. However, this measurement may have some issues. Perhaps the use of telephone and mobile use will not necessarily measure the desired effect of ICT diffusion, since they alone do not reflect any type of improvement in the form of communication.

4.2 Estimation method

According to Santos Silva and Tenreyro (2006), Ordinary Least Square (OLS) estimations under the presence of heteroscedasticity in the parameters will lead to biased estimates of true elasticities. One of the examples of biased estimation is the exaggeration of the effect distance and colonial ties in traditional OLS models. The authors propose a Pseudo Poisson Maximum-Likelihood (PPML) estimation method instead. Because the dependent variable is a count of the number of co-inventions that include a high number of results in zeroes with a highly skewed distribution, it is advisable to follow this estimation technique.

The dependent variable is calculated in levels as to avoid deleting observations that may in fact describe the nature of internationalization between countries, i.e. if an inventor in country A does not cooperate with an inventor in country B the value of that observation is zero. Thus, taking the natural logarithms of the dependent variable could bias the results by omitting the effect of the explanatory variables on the outcome. Count data models include the Poisson Pseudo-Maximum Likelihood (PPML) Estimator—the main estimation technique—, the Negative Binomial (NB) Model and the Zero Inflated Poisson (ZIP) Model. The former two have panel versions readily available in the econometric software of choice, while the ZIP model does not. However, there are suggestions to avoid this issue by pooling data and accounting for pairwise country clusters. As Herrera (2013) explains, every technique has advantages and disadvantages and it cannot be guaranteed that any estimation method outperforms the other. Thus, including several estimation methods has become common practice.

Additional model specifications have been suggested when in the presence of over-dispersion and a large number of zero values in the dependent variable. However, the effectiveness between the estimation methods, even when in presence of these two issues, still suffers an ongoing debate. It seems however, that the scale tips in favour of the Poisson Pseudo Maximum-Likelihood estimator, as it is claimed that this method still outperforms Negative Binomial and Zero Inflated estimation methods. The claim that a Negative Binomial estimation method would outperform Poisson would only be real in terms of efficiency, and when the nature of the over-dispersion is known, which is not the case in this thesis. When in the presence of datasets with a large numbers of zeroes, simulation evidence still supports PPML as the strongest performer between the three models (Santos Silva and Tenreyro, 2006; Shepherd, 2013). Furthermore, Santos Silva et al. (2014) develop a specification test that can be used to discriminate between non-negative count models using a simple-regression based specification test. After a series of assessments, it was determined that Poisson panel estimator and Poisson-Pseudo Likelihood estimation methods are appropriate for the recollected data.

Furthermore, although calculating the results with panel fixed effects method would control for observed and unobserved effects within the sample, it would omit all time invariant indicators because of collinearity with the fixed effects calculation. Since it is more interesting to see the measure and not just control for effects, fixed effects estimation in poison panel specification is not advisable. Although performing an interaction with the time variant indicators (i.e. the year dummies) is possible, it would make the estimation of results more difficult to interpret and requires significant computational power. Furthermore, random effects estimation for Poisson is not easily justifiable in this thesis and is not common in the literature because of the nature of the assumptions between the error term and the covariates. The PPML estimation method is preferred since it has been the one with the most presence in the reviewed literature and appears to be the most consistent with results in traditional gravity models of trade.

Furthermore, because of the nature of time series data and the variables selected for the analysis, it is possible to assume the presence of unit root across the sample. Be it the case and not accounted for could result in spurious results. However, if the variables show the characteristic of unit root but also show to be cointegrated, PPML estimation can continue.

5 Empirical Analysis

5.1 Descriptive Statistics

This section will provide the description for the variables of interest of the model. Starting from an overview of the data in order to understand the data in general terms via mean, standard deviation and value ranges for the continuous and categorical values. It will be followed by a detailed description of the dependent variable. It should be noted that the descriptive statistics take into account the sample reduction that was mentioned earlier, a list of the countries can be found in appendix 3.

The following table presents the summary statistics for all the continuous variables in levels. The subscript determines whether the variable relates to a single country[4] (i), country pair (ij) and if it is time variant (t) or invariant, i.e. the distance indicators are between country pairs (ij) but is time invariant, since it does not vary over time and has no subscript (t)[5]. The information does not include observations that correspond to the same country pairs $(i \neq j)$.

Table 1. Descriptive statistics: Continuous variables, adapted from OECD & CEPII

Variable	Obs.	Mean	Std. Dev.	Min	Max
$coinv_{ijt}$	45 120	8,35	47,02	0	993
pat_{it}	45 120	2344,05	6819,05	0	52332,37
$gdptot_{it}$	44 650	8,30E+11	1,86E+12	7,40E+09	1,41+E13
$poptot_{it}$	45 120	8,43E+07	2,38E+08	263725	1,35E+09
int_{it}	44 462	33,18	29,77	0,0002	96,21
tel_{it}	45 120	38,67	16,73	0,87	74,76
mob_{it}	45 120	60,02	47,44	0	172,32
$dist_{ij}$	45 120	5 804,85	5001,16	59,62	19586,18
$tdiff_{ij}$	45 120	3,85	3,54	0	12

Table 1 is divided in four tiers, each representing groupings of the indicators as they have been presented in the data description. The first tier includes the dependent variable ($coinv_{ijt}$), which shows a high degree of skewness[6] with a standard deviation more than a dozen times larger than the size of the mean. This reflects the high degree of heterogeneity in

[4] Since descriptive statistics do not vary between country of origin (i) or partner (j), only one is shown (i)
[5] Same country pairs are controlled for.
[6] It also shows over-dispersion, in which the variance is higher than mean by a large value

the international cooperation of patents, with the number ranging from zero to nine hundred and ninety-three (993) co-inventions.

The descriptive statistics for the attractors are included in the second tier in Table 1. All the variables show a high degree of variability between countries and time, with a relatively high standard deviation. This behaviour is expected since the database contains a very large set of countries with a wide range of innovative activities, economic and population characteristics.

The third tier in Table 1 includes the variables that will proxy for the diffusion of Information Communication Technologies (ICT). In their basic form, they measure diffusion of internet, telephone and mobile use and access, by percentage of population. Internet and mobile diffusion share a higher standard deviation than the mean across countries and time, which indicate a higher dispersion of access between the countries. Telephone access is more uniformly distributed across the sample, which may fall in line with expected values.

The fourth and last tier in Table 1 shows the descriptive statistics for the continuous, time invariant measurements of geographical distance between the country pairs. The simple distance between most populated cities between the countries averages at about 5 000 kilometres, with a minimum distance 59,62 kilometres and a maximum of around 19 000 kilometres. The time zone variable indicates that the average hour difference between countries is around four hours, ranging from zero to twelve (the maximum possible amount). This provides a sense of scope of the size of the sample of countries in the comparison. Including this large number of countries, dispersed around the globe is expected to have this range of latitude and longitudinal distance. From the results, it does not seem that the countries are agglomerated in space.

Tables 2 to 5 show the descriptive statistics for the categorical values. Since they represent time invariant variables that indicate relationships between the country pairs, the data is presented without the time dimension in order to properly reflect the characteristics of the sample. Thus, the observation size is "reduced" to 2 256, which reflects the total number of country pairs in the database, without counting same country pairs $(i \neq j)$.

Table 2. Descriptive statistics: Common border, adapted from OECD & CEPII

$contig_{ijt}$	Freq.	Percent
0	2 132	94,50 %
1	124	5,50 %
Total	2 256	100 %

Table 2 shows the variable for contiguity between the country pairs. As expected in a sample of this size, most country pairs do not share a common border, but it is likely that this will have a positive relationship with cooperation. As with the rest of the categorical variables in binary form, the reference will be not sharing a border. Thus, the effect will be measured on doing so, represented by a value of one (1).

Table 3. **Descriptive statistics: Common official language, adapted from OECD & CEPII**

$comlang\_off_{ijt}$	Freq.	Percent
0	2 118	93,88 %
1	138	6,12 %
Total	2 256	100 %

Table 3 shows the variable which measures sharing at least one common official language between country pairs. It is possible to see that around one in ten of the country pairs share an official language. It seems to account for a relatively large share of the country pairs, when compared to the rest of the categorical variables. Sharing a common language is expected to be positively related to increased cooperation between inventors.

Table 4. **Descriptive statistics: Common Language 9% population, adapted from OECD & CEPII**

$comlang\_ethno_{ijt}$	Freq.	Percent
0	2 086	92,46%
1	170	7,54%
Total	2 256	100%

Table 4 shows the variable that measures sharing a common language with at least 9 % of the population between country pairs. It is expected that this value is similar and somewhat larger than the official language variable. It is hoped that the variable may account for a second common language between countries, and may account for other shares of languages when they are not considered as official, but may affect the dependent variable.

Table 5. **Descriptive statistics: Shared colonial past, adapted from OECD & CEPII**

$colony_{ijt}$	Freq.	Percent
0	2 180	96,63%
1	76	3,37%
Total	2 256	100%

Table 5 shows the variable that measures if the country pairs have ever shared a colonial past. Interestingly, only a small fraction of the country pairs has ever been in a colonial relationship in this sample. As mentioned earlier, perhaps it may account for some cultural similarities between the country pairs and it is expected to be positively associated with the dependent variable.

In conclusion for the summary statistics, it is possible to see a high degree of over-dispersion in most of the continuous explanatory variables, with the exception of the fixed telephone line indicator and the geographical distance measurements. It seems that the data behaves in an expected manner since the sample includes information for many different countries and country pairs with heterogeneous characteristics. There is no significant loss of information due to missing data with this sample selection.

5.1.1 Correlation analysis

Because of the large sample size and the number of variables in the study, it is possible to foresee some issues with the model specification. For this reason, and to gain a better understanding of the data, a correlation analysis is included between all the variables in the model in their base level. It is also expected that some of the ICT indicators, the geographical variables and perhaps some of the attractors are correlated with each other. Thus, information from this analysis will help to anticipate some issues.

A Spearman correlation matrix[7] shows the relationship between the dependent and independent variables in the sample. At first instance, it can be seen that all of the explanatory variables and controls have the expected sign and non-zero values. The attractor variable for patents in the shows issues of high correlation with the variables for GDP, internet and mobile in both origin and partner countries. The variables for internet in origin and partner country are highly correlated with each other, suggesting a similar diffusion pattern of internet access between the country pairs, over the period. As expected, all the ICT indicators are correlated with each other by country of origin and partner country. The variables for mobile access between country of

[7] See Appendix 4. Spearman was used because of non-normal data in the sample.

origin and partner country show a high degree of correlation as well. As with the case of internet, it seems that mobile access increased in similar patterns between the country pairs. The geographical distance variables of kilometres and time difference between countries show a high degree of correlation. The two language variables also show high degrees of correlation.

The correlations between these variables have to be taken into consideration when analysing the results from the model. Some degree of correlation between the explanatory was expected, i.e. the ICT and geographical variables, and could be controlled for in the model specification through several methods. Furthermore, because omitting the attractors does not seem a viable option for the model specifications. The high correlation between them must be taken into consideration when interpreting the results from the model.

5.1.2 International cooperation in patents

Further considerations and descriptions are in order for the dependent variable. Due to its nature as a measurement of "flows" between country pairs, it is important to understand the special characteristics that this type of variable is usually associated with. Although these characteristics are expected, it is important to address them as potential limitations of this thesis. Furthermore, it is of interest to understand the behaviour of the dependent variable within a panel perspective, i.e. over time and between the countries of interest.

One special characteristic of the dependent variable in the study, and a frequent one in gravity type models, is the frequency of zeroes. As mentioned earlier, it is normal for count data, especially when measuring flows, to encounter a high degree of zeroes in the data. This information is valuable because it measures the absence of flows between countries, and should not be omitted since it may provide useful information for the model. That said, care should be taken when interpreting the estimates in this study, as in the case of this thesis, the frequency of zeroes is quite large.

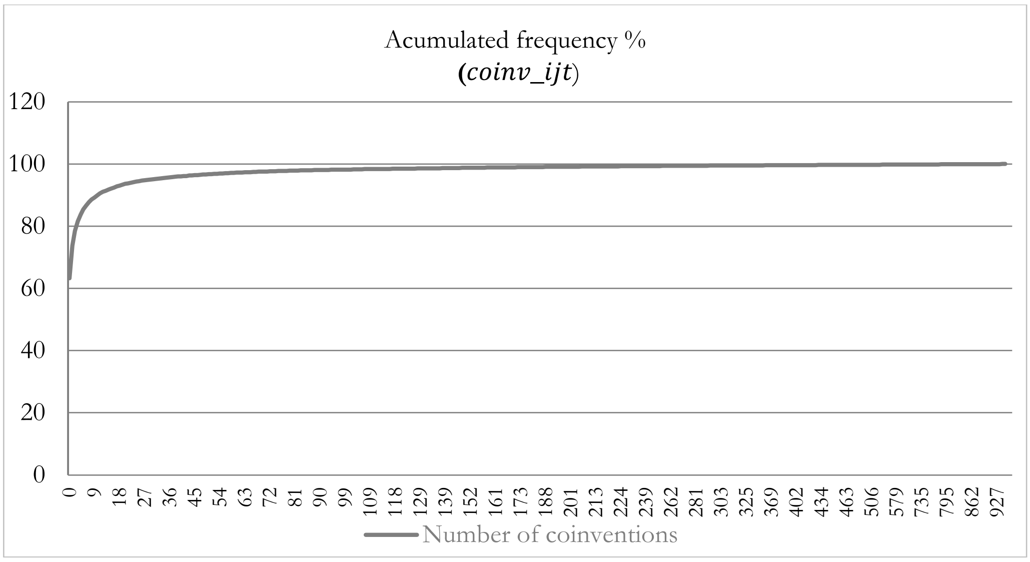


Figure 1. **Accumulated frequency, dependent variable, addapted from OECD**

Figure 1 shows the accumulated frequency of the variable in question. It shows the accumulated percentage for the number of appearances of each value in the data. It is possible to see the zero *inflated* characteristic of the dependent variable. There exists a very large number of zeroes in the database, accounting for around 63% of the total dispersion. It is also possible to see that the frequency increase behaves in an exponential manner. Although all of the values show information regarding flows, the high frequency of zeroes should be taken into consideration when selecting the method and when interpreting the results from the model.

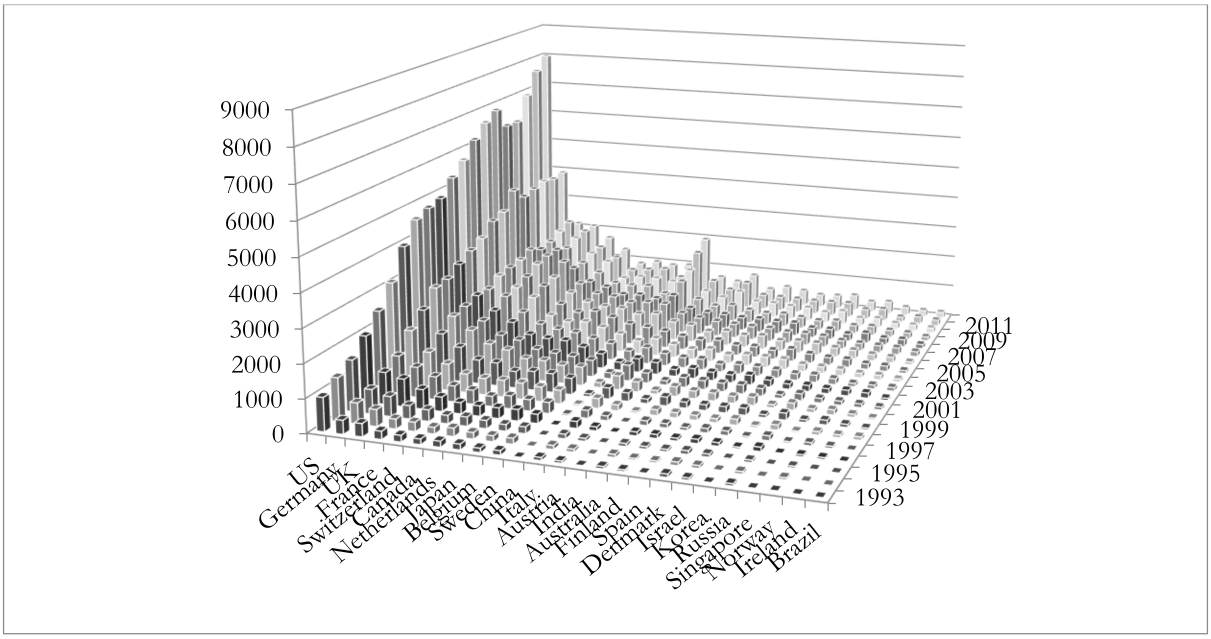


Figure 2. Total sum of co-inventions by country: Top 25 sample, adapted from OECD

Figure 2 shows a 3D graph, which is useful to see the interaction of the multiple dimensions of the database. The total sum of the number of co-inventions by the top 25 countries over the time period of analysis 1993-2012 is shown[8]. It should be noted that in countries like Ireland or Brazil—the last positions in this figure—are already relatively close to zero cooperation (as total sum per year), which also reflects the interpretation from figure 1. The omission of zero flows is not warranted in the model and this is done purely out of visual sake into the structure of the data. In other words, to better observe the behaviour of cooperation flows (over time) in the most significant of cases. Observing the data for international cooperation in patents in this form gives crucial perspective in how the data is distributed between the countries on the time perspective. It is possible to see that only a handful of countries have a relatively significant number of international cooperation across the period. However, an important characteristic that should be noted is the overall increase of the size of the indicator over time. Wherever there is cooperation, it seems to be incremented in the following periods.

The total sample sum of patent cooperation per country might seem underwhelming at first, especially the figure with the full sample. It should be prudent to remember however, that this takes into account all the non-cooperating countries in the sample. The maximum number of co-inventions is 8 001 for the United States in the year 2012.

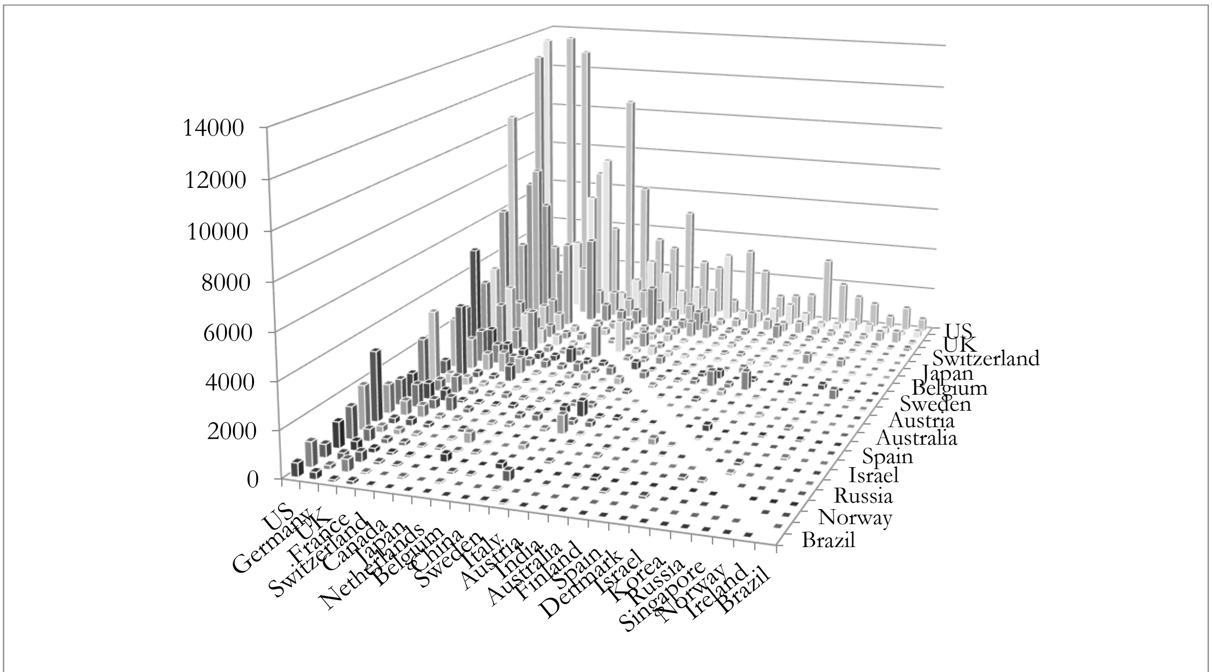


Figure 3. **Total sum of co-inventions by country pair: Top 25 sample, adapted from OECD**

[8] See appendix 5 for original database of 92 countries and period 1990-2012

Figure 3 shows the total sum of cooperation in patents by country pair. As explained before, most of the sample contains data close to zero[9]. The country pairs with the highest level of cooperation across the time period are, amongst others, the United States with the following countries in descending order: Germany, the UK and Canada; followed by the pair of Germany and Switzerland and the United States and Japan.

Table 6. **Top ten country pair patent co-inventions, adapted from OECD**

Country pair		Year	Co-inventions
United States	Germany	2012	993
United States	Germany	2006	955
United States	Germany	2011	951
United States	China	2012	927
United States	United Kingdom	2007	919
United States	Germany	2010	918
United States	Germany	2007	916
Germany	Switzerland	2007	903
United States	United Kingdom	2011	896
United States	Germany	2005	894

Table 6 shows the top ten country pairs with the highest amount of cooperation in the database, regardless of year. It is interesting, although perhaps not unexpected, that the top positions are dominated by the United States and Germany, and range from 2005 to 2012. Nevertheless, it is interesting to see the United Kingdom, Switzerland and China in the top ten country pairs in cooperation. Although not the main focus of the thesis, it should be noted that the inclusion of China in the top ten co-inventor countries is quite noteworthy. It is part of the second highest cooperation in the year 2012, the first one being between the United States and Germany, and somehow seems to indicate that the inventiveness of the country is important enough to attract the attention of the biggest innovator (in absolute terms) in the database.

5.2 Results

Prior to the estimation of results, unit root and cointegration tests are performed to determine its presence amongst the time variant variables. As mentioned earlier, not all the tests can be performed on the original database of 91 countries and time period 1990-2012 because of missing data and gaps in the database; only unit Fisher type unit roots tests could be performed in some cases, with the exception of the variable for co-invention, because it allows for

[9] See Appendix 6 for figure with complete set of countries

unbalanced data with gaps. Furthermore, calculating the presence of cointegration on the variables is not possible for the same reason. No matter how the data was entered, the cointegration test could not be performed in either the Westerlund or the Pedroni methods[10]. As mentioned earlier, this issue led to a decision to reduce the database in order to perform the tests[11].

Table 7. Unit root test

		Levels		1st Difference	
Variable	Test	Constant	Constant & Trend	Constant	Constant & Trend
coinv	HT	0,9181	0,7435	-0.1570***	-0.2517***
	Breitung	-30.118***	-26.7172***	-1.2e+02***	-59.3342***
	Fisher*	-32.1004***	-38.8325***	-147.8968***	-125.1915***
lnpat_o	HT	0.8126***	0,9805	-0.1605***	0.0361***
	Breitung	49,3993	26,1035	-1.1e+02***	-56.0940***
	Fisher	-59.4965***	4,0192	-147.8968***	-125.1915***
lngdptotom	HT	0,9601	0,9979	0.3328***	0.4515***
	Breitung	95,2028	47,5033	-73.9868***	-57.8127***
	Fisher	10,2258	25,846	-62.8459***	-56.2177***
lnpoptotom	HT	0,9877	1,0005	0.3616***	0,8686
	Breitung	135,4691	98,1069	37,4115	38,0823
	Fisher	25,8059	-36.4791***	-20.8536***	-51.0136***
int_o1	HT	n/a	n/a	n/a	n/a
	Breitung	n/a	n/a	n/a	n/a
	Fisher	56,1978	28,3597	-21.9540***	-4.86***
tel_o1	HT	0,9411	0,981	0.3214***	0.3981***
	Breitung	23,2848	56,9853	-42.4786***	-55.5015***
	Fisher	3,3967	21,3545	-16.5950***	-23.5694***
mob_o1	HT	0,9812	0,9893	0.5472***	0.5607***
	Breitung	103,8641	58,9504	-42.2871***	-20.3654***
	Fisher	15,8026	5,0944	-37.4184***	-18.1662***

*Fisher reports Inverse normal Z

Table 7 shows the results for unit root testing for HT, Breitung and Fisher type tests, which were selected because they are more appropriate for the large N in the database. The results show that the dependent variable has relatively weak evidence of unit root and I(1). The patent, GDP and

[10] In addition, calculating each unit root test could take up to 12 hours of computing time. Which led to extremely slow progress.
[11] See appendix 3 for final list of countries in the model, time period is also reduced to 1993-2012

Population indicators in natural logarithms show stronger evidence for unit root and I(1), as do the indicators for ICT; Internet, telephone and mobile access. Due to the results, further testing for cointegration can continue.

Table 8. Westerlund cointegration test (a)

Statistic	Value	Z-value	p
Gt	-4,216	-95,178	0,0000
Ga	-3,20E+13	-2,00E+14	0,0000
Pt	-91,356	-0,854	0,1970
Pa	-3,279	28,687	1,0000

H0: No cointegration
2304 series and 5 covariates

Table 8 shows the result for the cointegration test designed by Persyn & Westerlund (2008). The results show evidence that the indicators are co-integrated in the long run in at least some of the panels. However, the variable for internet diffusion (int_o1) could not be calculated in this model because of the gaps in the series.

Table 9. Westerlund cointegration test (b)

Statistic	Value	Z-value	p
Gt	-4,153	-77,859	0,0000
Ga	-2.4e+13	-1.3e+14	0,0000
Pt	-89,44	6,986	1,0000
Pa	-2,636	38,726	1,0000

Results for H0: no cointegration
With 2116 series and 6 covariates

An additional test was performed with the necessary sample reductions to ensure that the calculation was possible, more specifically the omission of the country Malta. This additional calculation also shows evidence of co-integration for at least some of the panels. Although because of the reduction of observations, its interpretation should be handled with some care. These tests show at the very least a possible long-run relationship between the variables that may help better explain the results of the following section[12]. The results from these tests appear to

[12] Previous gravity literature that acknowledges the issue of unit root and co-integration usually performs the estimation with OLS or equivalent lineal interpretations. No examples of PPML or other Poisson-type distributions were found.

indicate that it is possible to proceed with the calculation of the results since the variables appear to be cointegrated in at least some of the panels.

The results will show the calculations for the aforementioned model specifications, with the inclusion of two additional robustness models, which will be explained accordingly. As mentioned previously, all models are calculated with Poisson Pseudo-Maximum Likelihood and include country dummies *i* and *j* with time dummies *t* to proxy for observable country and time effects in the sample. All models are calculated with clustering on country pairs to control for autocorrelation between countries and include robust standard errors to anticipate heteroscedasticity. By default, PPML allows the option to exclude problematic observations from the estimation to ensure the calculation of the results[13].

Table 7[14] shows the results for the different estimation outputs for the equations 4 to 9, with some additions[15]. **The non-augmented model (4)** shows the estimation output for the first model specification. It shows that the attractors for Patents and GDP, and the distance measured in kilometres are statistically significant in the model. The signs are as expected for the patent and GDP variables, indicating that richer and more inventive countries will likely attract cooperation at the inventor level.

The economic significance of the coefficients also seems promising, indicating that the variables are likely to have a meaningful relationship with the dependent variable. In this case, the coefficient for patents indicate can be measured as elasticities and state that, holding everything else constant, the increase of patent applications by 1 unit of (ln) patents increases logs of expected counts for patent co-invention by its coefficient of 0,351. In other words, an increase by a factor of 1,42[16] (increase of 42,05%). The sign of the population attractor seems to indicate that there is a negative association between country size and cooperation between inventors across countries, which seems to be reflected by a supply effect, in which local economic agents seek less workers and expertise abroad (Montobbio & Sterzi, 2012).

---

[13] Only observations related to country and time effects are dropped in this manner

[14] Shown below for spacing

[15] As mentioned earlier, this is a smaller sample than originally intended. However, the outputs are quite similar even with just 1/4 of the original sample. This may suggest that the subsample shows robustness, or that the selected country pairs (which were the ones with highest count of co-invention) drive the results. The results from the larger sample are included in appendix 7.

[16] Calcultated by $e^{0,351}$

Table 10. **Estimation results for International patent cooperation**

Exp. variables	**Eq. 4 Non-Aug**	**Aux1 Aug**	**Eq. 5 Aug ICTLvls**	**Eq. 6 Aug ICTTrans**	**Eq. 7 Aug Int**	**Eq. 8 Aug ICTavg**	**Aux2 Aug LnICTavg**
lnpat_it	0.351***	0.361***	0.369***	0.366***	0.358***	0.361***	0.358***
	(0.0764)	(0.0770)	(0.0798)	(0.0805)	(0.0769)	(0.0793)	(0.0804)
lnpat_jt	0.351***	0.361***	0.369***	0.366***	0.358***	0.361***	0.358***
	(0.0764)	(0.0770)	(0.0798)	(0.0805)	(0.0769)	(0.0793)	(0.0804)
lngdptot_it	0.810***	0.795***	0.760***	0.737***	0.792***	0.773***	0.782***
	(0.235)	(0.236)	(0.223)	(0.258)	(0.236)	(0.249)	(0.249)
lngdptot_jt	0.810***	0.795***	0.760***	0.737***	0.792***	0.773***	0.782***
	(0.235)	(0.236)	(0.223)	(0.258)	(0.236)	(0.249)	(0.249)
lnpoptot_it	-0.383	-0.224	-0.0628	-0.171	-0.231	-0.236	-0.258
	(0.333)	(0.332)	(0.349)	(0.325)	(0.333)	(0.334)	(0.341)
lnpoptot_jt	-0.383	-0.224	-0.0628	-0.171	-0.231	-0.236	-0.258
	(0.333)	(0.332)	(0.349)	(0.325)	(0.333)	(0.334)	(0.341)
internet_it			0.0346				
			(0.0979)				
internet_jt			0.0346				
			(0.0979)				
telephone_it			-0.0438				
			(0.194)				
telephone_jt			-0.0438				
			(0.194)				
mobile_it			0.0892				
			(0.0774)				
mobile_jt			0.0892				
			(0.0774)				
dint_ijt				0.151	0.144		
				(0.110)	(0.116)		
dtel_ijt				0.0867			
				(0.183)			
dmob_ijt				-0.315***			
				(0.0886)			
avgdICT_ijt						-0.183	
						(0.216)	
lnavgdICT_ijt							-0.0454
							(0.0380)
lndist_ij	-0.455***	-0.185***	-0.178***	-0.159***	-0.176***	-0.176***	-0.173***
	(0.0308)	(0.0576)	(0.0569)	(0.0576)	(0.0573)	(0.0568)	(0.0560)
tdiff_ij		-0.0147	-0.0174	-0.0156	-0.0180	-0.0167	-0.0165
		(0.0143)	(0.0142)	(0.0140)	(0.0142)	(0.0143)	(0.0143)
contig_ij		0.425***	0.422***	0.438***	0.432***	0.416***	0.413***
		(0.0816)	(0.0811)	(0.0821)	(0.0815)	(0.0816)	(0.0825)
comlang_off_ij		0.672***	0.684***	0.688***	0.684***	0.685***	0.686***
		(0.111)	(0.110)	(0.107)	(0.110)	(0.109)	(0.109)
comlang_ethno_ij		0.0192	0.0124	0.0194	0.00688	0.0178	0.0201
		(0.110)	(0.109)	(0.107)	(0.110)	(0.109)	(0.109)
colony_ij		-0.0414	-0.0499	-0.0684	-0.0503	-0.0534	-0.0547
		(0.0611)	(0.0612)	(0.0591)	(0.0609)	(0.0602)	(0.0597)
Constant	-11.17***	-15.34***	-15.81***	-15.04***	-15.27***	-15.00***	-15.21***
	(2.293)	(2.157)	(2.486)	(2.512)	(2.142)	(2.308)	(2.244)
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44,136	44,136	43,228	43,228	43,228	43,228	43,228
R-squared	0.899	0.955	0.955	0.958	0.955	0.955	0.956

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Numbers based on model specification

However, the coefficient is not statistically significant. The sign is also expected for the distance indicator, which shows a negative association between geographical distance and cooperation between inventors. The coefficient of -0,455 shows that, holding everything else constant, an increase of distance between countries by 1 unit of (ln) distance[17] results in a change of cooperation by a factor of 63,44% (decrease of 36,55%).

Appendix 8 shows a margin analysis for (ln) distance and the corresponding number of co-inventions across the sample between minimum and maximum distances between the country pairs, in the x-axis. The figures are calculated in kilometres calculating the exponential values for one-unit increases in (ln) distance. It shows an exponential decrease over distance, reflecting the results from the model. The value of the coefficient can be seen in the exponential equation for the fitted line (an approximation of the actual value).

In order to provide an accurate measurement of the role of ICT in the comparisons between model specifications, an additional **augmented model with controls (aux1)** is calculated. The signs of all coefficients are equal in terms of the previous model (4) with some minor changes in their size. The coefficients for the patent attractor are slightly larger than in the previous model, although the increase is marginal and an interpretation may not necessarily hold for this change. The coefficients for GDP and Population decrease in absolute size. The coefficients for time difference, contiguity, the common language dummies and colonial ties appear to contain the correct signs. However, the results for time difference, common language by 9% of the population and colonial ties are not statistically significant. The relationship with contiguity seems to be quite strong with the highest sized coefficient for geographical and cultural ties in the results of this model. According to the coefficient, sharing a common border with another inventor increases the magnitude of the relationship by a factor of around 153% (increase of 52,95%) when compared to not sharing a border with another inventor, holding everything else constant. The most interesting aspect of the model is the reduction in the coefficient of (ln) distance. It seems that controlling for other measurements of distance reduces the magnitude of the relationship of geographical distance to a factor of 83,11% per increase of 1 unit of (ln) distance, which seems to follow the *a priori* expectations and previous research in the joint relationship of distance and bilateral interactions.

Appendix 9 shows the margin analysis when controlling for the rest of the distance and cultural measurements. Comparing the results with Figure 4 shows that the decrease is lower in magnitude. It seems that distance *per se,* measured as simple distance between countries, may be

[17] Appendix 8 reflects the values for ln(distance)

overestimated if controls are not included that may account for other types of distance measurements in the sample.

The augmented model with ICT measurements in levels (5) shows the estimation output, which includes ICT indicators for internet, telephone and mobile access in, levels per inhabitant, for country *it* and for country *jt*[18]. In terms of expected signs and statistical significance, there are only two differences relative to the previous model (aux1). The coefficients of the attractor for patent applications show that the relationship is larger when including the variables for ICT, although the difference seems moderate. The coefficient sizes for geographical distance do not seem to suffer much change either but do marginally decrease in size. However, the ICT indicators themselves do not appear to be statistically significant and the coefficients seem to be small in terms of economic significance, indicating an increase in patent cooperation of 3,46% with an increase of 1 internet user per inhabitants across country pairs, ceteris paribus. Furthermore, the negative coefficients for telephone access may not necessarily reflect expected relationships. However, this was somewhat expected since the ICT indicators show a high degree of multicollinearity.

The augmented model with ICT diffusion (6) shows the estimation output that contains the initial transformation for the ICT variables, which aims to proxy diffusion. The transformation aims to calculate the difference in internet, telephone and mobile diffusion between country pairs *ijt,* per inhabitant. When comparing the results with model 5, it seems that most of the coefficients contain similar results, with some exceptions. The coefficient for geographical distance in kilometres seems to be relatively smaller than the rest of the model specifications, although the size reduction is marginal. Cultural ties seem to follow similar size and signs. The ICT indicators are not statistically significant in the model and incorrect in sign, with the exception of the difference between mobile telephone diffusion between country pairs. As mentioned earlier, increasing said distance may result in a negative association with international cooperation as it may reflect similar patterns of ICT access. However, the signs for internet users and telephone access do not reflect this intuition in the model. Nevertheless, the indicators are highly correlated with each other, which may be the cause for this reversion of sign, in a similar fashion to model five (5).

The augmented model with the diffusion of Internet (7) shows the estimation output when only including one of the ICT indicators in its estimation of difference in diffusion, the internet. Most of the coefficients and signs are comparable to the previous models 5 and 6. However, the

[18] Instead of percentage, for rescaling purposes since PPML is sensitive to this issue.

coefficient for difference in internet use remains statistically significant and still shows an unexpected sign, which would indicate an association of increases in co-invention when the difference in diffusion patterns between country pairs increases.

The augmented model with average ICT diffusion (8) shows the estimation output for the calculation of average difference between the ICT indicators between countries. This attempts to overcome the possible issues from omitting some of the ICT indicators. Again, most of the coefficients show similar signs and size across the previously stated models, paying special attention to the geographical indicators. The coefficient for average ICT appears to be not statistically significant. The coefficient shows a negative association with the dependent variable with a decrease in patent cooperation by a factor of 83,28% (decrease of 16,72%) per unit increase of average difference between ICT diffusion between country pairs.

Finally, **the augmented model with ln average ICT diffusion (aux2)** aims to provide an additional measurement for the ICT variables in average terms. This time, the natural logarithm is calculated since it would not result in loss of information and may address some issues in modelling. It is possible to see that most of the coefficients and signs remain similar to previous models. The coefficient measurement of (ln) average ICT diffusion difference between countries shows a negative sign and is not statistically significant. Thus, there does not seem to be a statistical relationship with the average calculation of the difference of ICT diffusion.

5.3 Discussion

Overall, it is possible to see from the estimation outputs of all the models that the indicators related to the attractors appear to be robust across estimations, since they do not seem to vary much with the different model specification, other than some marginal increase and decrease in coefficient size. However, it should be noted that there was a high correlation between these variables, which may produce unseen errors in the estimations. It also is possible to see similarities across the geographical and cultural indicators across the augmented models, without suffering much change with the inclusion of the different measurements for ICT, in levels, measured as proxies for diffusion by technology, and as proxy for average ICT diffusion.

Language barriers seem to be an important factor, as evidenced by the statistical and economic significance of common official language between countries, a result partly shared with Picci (2010) and by Montobbio & Sterzi (2013). Furthermore, the latter study also finds that when including the language indicator, colonial past is not statistically significant. These results are shared in this thesis as well, although the estimation technique is the same (PPML). Interestingly, this result appears to be replicated in the use of PPML in gravity equations for trade theory as well.

On the role of distance between country pairs, it is possible to see additional similarities with Montobbio & Sterzi (2013). The coefficients of the non-augmented and augmented model in this thesis appear to be of relatively similar size, even when the choice of attractors differs somewhat and the role of ICT is not included in their analysis, and with the use of different control variables like Trade, FDI and Intellectual Property Rights. In the study by Picci (2010), the role of distance in the presence of all controls seems to be slightly overestimated when compared to the ones mentioned earlier. This appears to be in line with the argument from Santos Silva & Tenreyro (2006), in which they claim that OLS in the presence of heteroscedasticity will lead to an overestimation of this indicator.

One important aspect to keep in mind is that not all patenting collaborations will be of similar value. This is a result of a typical skewness in the technological and economic value of inventions. However, robustness checks for patent value by Montobbio & Sterzi (2013) find that when controlling for forward citations in the next three years by patent collaboration, find no statistical significance for the measurements of distance or time zone difference. This robustness check indicates that if a measure of the value of the patent collaboration is included, geographical distance does not play a role in collaboration as long as the innovation is worthwhile. In the words of the authors, "the effects of transportation and communication costs are negligible" (Montobbio & Sterzi, 2013: pp. 295). This additional step could not be replicated in the thesis due to a lack of data regarding forward citations. This has to be seen as a possible limitation since the quality of the cooperation is not controlled for.

Different estimations for ICT do not seem to provide much change in terms of the role of distance with co-invention between individuals, and the possible presence of multicollinearity between the three indicators does not provide any advantages either. There is not a clear model of interest between the model specifications. While all seem to have a quite high goodness of fit, there is no additional information provided after including the different measurements for distance and cultural ties in the first augmented model. It seems that most of the co-inventions are explained by the augmented model without ICT indicators, presented by the auxiliary model one (aux1). It seems plausible that these models may in fact be representative of the actual environment surrounding tacit knowledge flows for the period.

If there was an overall significant relationship between ICT and tacit knowledge, the period for the study, may not necessarily reflect the advantages of advanced technology in terms of ICT even though it includes years up until 2012. There is, of course, a possibility that a relationship exists, but was not able to be identified with the model specification, estimation method and/or data. Nevertheless, if there was an actual influence, it is not likely that it will completely negate the effect of geographical distance for tacit knowledge at its present state and even less likely, be reflected in the period.

There is however, a drawback from this selection of variables if the intent is to measure their relationship in an international scope. Although accounted for as official ICTs in databases, they may not necessarily reflect communication between countries. Indeed, it seems more likely that individuals will communicate with each other via the internet than with mobile and telephone calls. Since the latter two are usually costlier to perform and are more limited in the information that can be communicated.

The influence of the internet in all aspects of life rapidly changes over time and the advancements in the technology from one year to the other can vary even more so. The current period ends in 2012, which means that precious years in this technological advance will not be taken into account. However, even if the data was available, it may still not be reflected in a study of this type since this influence may just only be starting to have an initial echo in the data.

Unfortunately, quality and the improvement of the communication are not measured in the thesis. Perhaps it is not the diffusion of the service *per se,* that allows for better remote communication, but the speed and quality of the services, mainly the internet service[19]. This cannot be reflected at all in the average measurement of the ICT indicators or its transformations to proxy diffusion. The measurement of internet access is limited since it is only a proxy for diffusion, but not the quality of the service. Countries with high quality services will not be separated from those of low quality and they are not expected to evolve equally over time. This is a main limitation

ICT may very well play a role in cooperation between inventors across countries. However, the role may not be large enough to be reflected in the data and to negate the necessity for face-to-face communication in the flow of tacit knowledge. Perhaps this will be overcome with the advancement of technology, but this remains to be seen. For the moment, the theoretical argument in which ICTs are seen as inadequate to communicate tacit knowledge across large distances seems to hold, due to inventors preferring face-to-face communication.

6 Conclusion

The thesis aimed to provide some insight into the relationship of Information Communication Technologies and geographical distance in the context of tacit knowledge flows. Traditionally

[19] …since it seems that most of the improvements in technology are for this service, and perhaps mobile phones.

and in a theoretical perspective, tacit knowledge flows are not expected to be influenced by improving ICTs. More precisely, geographical distance is robust to improvements in communication technology because of the nature of the transmitted knowledge. Face-to-face communication is required in order for the flow to be successful in communicating essential aspects that are needed for a worthy use of the knowledge.

The objective was to challenge this notion with an empirical approach, which seems to be lacking a clear structure along the literature. There are several considerations when it comes to the theoretical approach or modelling, the choice of data to account for the possible attractors and distances, and the estimation methods. Furthermore, the empirical application of gravity type models to tacit knowledge flows seems to be in its infancy, at its current state. Hence, there is much room for improvement and refinement in this sense.

Considering the previous discussion of the results of the analysis, the following can be said about the hypotheses of the thesis: Firstly, on the relationship between geographical distance and tacit knowledge flows, there is enough evidence to reject the null hypothesis of no negative effect of geographical distance in tacit knowledge flows, with patent co-invention as proxy. There appears to be a clear influence of distance in the likelihood of an interaction between inventor and co-inventor across different countries. Secondly, on the effect of the diffusion of Information Communication Technology on the relationship between geographical distance and tacit knowledge flows, there is not enough thorough evidence to reject the null hypothesis of no negative effect in the association between geographical distance and tacit knowledge flows. Although there is a slight decrease on the coefficient size of the distance indicator between country pairs, it does not seem to be sufficient change to reject the hypothesis. There might be an effect, but it is not large enough in this context to conclude otherwise. Finally, on the effect of the diffusion of (only) the influence of Internet on the relationship between geographical distance and tacit knowledge flows, there does not seem to be enough evidence to reject the null hypothesis in a similar fashion to the previous measurement. There is not sufficient evidence to indicate that the internet closes the gap left by geographical distance when inventors and co-inventors choose to cooperate on an international level.

Following the segmentation of each of the hypotheses' rejections, it is possible to provide an answer to the research question. With the data, specification, estimation methods and considerations taken in the thesis, there is not sufficient evidence to indicate that Information Communication Technologies negate the effect of geographical distance on tacit knowledge flows at the inventor level, measured as international co-invention. This ultimately falls in line with the theoretical context surrounding the topic as expressed in the Theory section of the thesis.

All the evidence gathered in this thesis seems to indicate that the role of geographical distance in tacit knowledge flows is not dwindling. Previous research shows general stoicism across different types of study, not only for tacit knowledge, but also for other types of flows in an international scope. Across the sample and the selected period in the analysis, geographical distance did not suffer much change with the extensive choice of indicators to account for an influence of Information Communication Technologies. This would suggest merit to the theoretical outline of tacit knowledge. Indeed, it seems that this relationship between distance and flows will endure for the time being. At the very least, this is what may be anticipated from the evidence gathered in this thesis.

A possible recommendation for future research is related to one main limitation in this study. It may be possible to include the quality of internet service in the analysis. The intuition being that only assessing access is not enough to tackle the issues in the difficult transfer of tacit knowledge. It is unlikely that low speed internet would be sufficient to replace the interactions that are usually reserved to face-to-face communication. However, perhaps it is only a matter of time before this effect is reflected in tacit knowledge flows. Improvements in technology always allow for innovative ways to deal with day to day issues, and perhaps this one is just a few year—or decades—away to be resolved.

Regarding practical implications of the thesis, it may be possible to argue for different and perhaps conflicting approaches. On the one hand, the importance of geographical proximity could be supported and efforts to agglomerate actors in the anticipation that this will promote that production of new knowledge, in the tacit form. Supporting policy that promotes free movement and grouping of creative efforts so that cooperation can be done face-to-face in order to increase inventive outputs in countries where this movement is possible. Increase cooperation with partner countries in an attempt to promote each other in terms of inventive prowess. On the other hand, it could be seen as a challenge to increase the possibility of long distance communication in a way that does not limit the success of transmitting tacit knowledge. Perhaps this will be a consequence of the inevitable improvement of ICTs. But perhaps a more directed effort could take place in order to improve the possibility of cooperation between inventors across great distances, in order to search for a potentially untapped knowledge reservoir which may reveal otherwise impossible possibilities.

7 References

Anderson, J.E., (1979). A Theoretical Foundation for the Gravity Equation. The American Economic Review, Vol. 69, No. 1, Pp.106-116.

Anderson, J.E., (2010). The Gravity Model, Working Paper 16576. NBER Working Paper Series.

Asheim, B., Coenen, L., Vang, J., (2005). Face-to-Face, Buzz and Knowledge Bases: Socio-spatial Implications for Learning and Innovation Policy. Paper no. 2005/14, CIRCLE, Lund University.

Ejermo, O. & Karlsson, C. (2006). Interregional Inventor Networks as Studied by Patent Coinventorships. Research Policy, No. 35, Pp. 412-430.

Head, K. & T. Mayer, (2013). Gravity Equations: Toolkit, Cookbook, Workhorse. Handbook of International Economics, Vol. 4,eds. Gopinath, Helpman, and Rogoff, Elsevier.

Herrera, E.G.j (2013). Comparing Alternative Methods to Estimate Gravity Models for Bilateral Trade. Empirical Economics, Vol. 44, Pp. 1087-1111.

Jaffe, A.B., (1988). Demand and Supply Influences in R&D Intensity and Productivity Growth. The Review of Economics and Statistics, Pp. 431-437.

Jaffe, A.B., Trajtenberg, M. & Henderson, R., (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. the Quarterly journal of Economics, Pp. 577-598.

Lucas, R.E., (1988). On the Mechanics of Economic Development. Journal of monetary economics, 22(1), Pp. 3-42.

Malecki, E.J., (2010). Everywhere? The Geography of Knowledge. Journal of Regional Science, Vol. 50, No.1, Pp. 493-513.

Maurseth, P.B. & Verspagen, B., (2002). Knowledge Spillovers in Europe: A Patent Citations Analysis. The Scandinavian journal of economics, Vol. 104, No. 4, Pp. 531-545.

Mayer, T. & Zignago, S., (2011). Notes on CEPII's distances measures: the GeoDist Database CEPII Working Paper 2011-25.

Montobbio, F. & Sterzi, V., (2013). The Globalization of Technology in Emerging Markets: A Gravity Model on the Determinants of International Patent Collaborations. World Development, Vol. 44, Pp. 281-299.

Moodysson, J., (2008). Principles and Practices of Knowledge Creation: On the Organization of “Buzz” and “Pipelines” in Life Science Communities. Economic Geography, Vol. 84, No.4, Pp. 449-469.

OECD (2001). Education Policy Analysis. OECD Publishing, Pp. 152.

OECD (2009), OECD Patent Statistics Manual. OECD Publishing, ISBN 978-92-64-05412-7.

OECD (2016). Indicators of International Co-operation. OECD Patent Statistics (database). DOI: http://dx.doi.org/10.1787/data-00507-en (Accesed on May 3 2016).

OECD (2016). Patents by Main Technology and by International Patent Classification (IPC), OECD Patent Statistics (database). DOI: http://dx.doi.org/10.1787/data-00508-en (Accesed on 1 May 2016).

Peri, G., (2005). Determinants of Knowledge Flows and their Effect on Innovation. The Review of Economics and Statistics, Vol. 87, No. 2, Pp. 308-322.

Persyn, D. & Westerlund, J., (2008). Error-Correction-Based Cointegration Tests for Panel Data. Stata Journal, Vol. 8, No. 2, Pp. 232-241.

Picci, L., (2010). The Internationalization of Inventive Activity: A Gravity Model Using Patent Data. Research Policy, Vol. 39, Pp. 1070-1081

Reinert, K.A., Rajan, R.S., & Glass, A.J. & Davis, L.S.J. (2010), The Princeton Encyclopedia of the World Economy (Two Volume Set), Princeton University Press

Romer, P.M., (1990). Endogenous Technological Change. Journal of Political Economy. Vol. 98, No. 5, pt. 2, Pp. 71-102.

Samuelson, L., (2004). Modelling Knowledge in Economic Analysis. Journal of Economic Literature, Vol. 42, No. 2, Pp. 367-403.

Santos Silva, J.M.C., Tenreyro, S. & Windmeijer, F., (2014). Testing Competing Models for Non-negative Data with Many Zeroes. Journal of Econometric Methods, Vol. 4, No. 1, Pp. 29-46

Seki, Ý., (2008). The Importance of ICT for the Knowledge Economy: A Total Factor Productivity Analysis for Selected OECD Countries. In Papers of the Annual IUE-SUNY Cortland Conference in Economics (pp. 72-90). Izmir University of Economics.

Silva, J.S. & Tenreyro, S., (2006). The log of gravity. The Review of Economics and statistics, Vol. 88, No. 4, Pp. 641-658.

Shepherd, B., (2013). The Gravity Model of International Trade: A User Guide. ARTNeT

Gravity Modeling Initiative, United Nations.

Solow, R.M., (1956). A Contribution to the Theory of Economic Growth. The Quarterly Journal of Economics, Vol. 70, No. 1, Pp. 65-94.

Steinmueller, W.E., 2000. Will New Information and Communication Technologies Improve the 'Codification' of Knowledge?. Industrial and corporate change, No. 9, Vol. 2, Pp. 361-376.

United Nations (2016). UN data. Avaiable at: http://data.un.org/Explorer.aspx?d=ITU ITU Website: http://www.itu.int/en/ITU-D/Statistics/Pages/definitions/default.aspx (Accesed May 10 2016).

WIPO (2016). PCT-The Ineternational Patent System. World Intellectual Property Organization. At: http://www.wipo.int/pct/en/faqs/faqs.html (Accesed on May 5 2016)

The World Bank (2016). World Development Indicators. Available at: http://data.worldbank.org/data-catalog/world-development-indicators (Accesed on April 27 2016)

The World Factbook 2016-17 (2016). Washington, DC: Central Intelligence Agency

World Trade Organization (2012). A Practical Guide to Trade Policy Analysis, WTO ISBN 978-92-870-3812-8

8 Appendix

Appendix 1. Full list of countries in the sample

Algeria
Andorra
Argentina
Armenia
Australia
Austria
Belarus
Belgium
Bermuda
Bosnia and Herzegovina
Brazil
Bulgaria
Canada
Cayman Islands
Chile
China (People's Republic of)
Colombia
Costa Rica
Croatia
Cuba
Cyprus
Czech Republic
Denmark
Djibouti
Ecuador
Egypt
El Salvador
Estonia
Finland
France
Georgia
Germany
Greece
Guatemala
Hungary
Iceland
India
Indonesia
Iran
Ireland
Israel
Italy
Jamaica
Japan
Jordan
Kazakhstan
Kenya
Korea
Kuwait
Latvia
Lebanon
Lithuania
Luxembourg
Malaysia
Malta
Mexico
Mongolia
Morocco
Netherlands
New Zealand
Nigeria
Norway
Pakistan
Panama
Peru
Philippines
Poland
Portugal
Puerto Rico
Russia
Saudi Arabia
Seychelles
Singapore
Slovak Republic
Slovenia
South Africa
Spain
Sri Lanka
Sweden
Switzerland
Thailand
Trinidad and Tobago
Tunisia
Turkey
Ukraine
United Arab Emirates
United Kingdom
United States
Uruguay
Uzbekistan
Venezuela
Zimbabwe

Appendix 2. Model specifications: full list

A.3.1. Augmented model with controls (aux1)

$$\begin{aligned} coinv_{ijt} = {} & \beta_0 + \beta_1 \ln(pat_{it}) + \beta_2 \ln(pat_t) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) \\ & + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8 \ln(dist_{ij}) \\ & + \beta_9(contig_{ij}) + \beta_{10}(tdiff_{ij}) + \beta_{11}\left(comlang_{off_{ij}}\right) \\ & + \beta_{12}\left(comlang_{ethno_{ij}}\right) + \beta_{13}(colony_{ij}) \end{aligned}$$

A.3.2. Augmented model with ICT Diffusion (eq. 6)

$$\begin{aligned} coinv_{ijt} = {} & \beta_0 + \beta_1 \ln(pat_{it}) + \beta_2 \ln(pat_t) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) \\ & + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8(dint_{ijt}) \\ & + \beta_9 ln(dtel_{ijt}) + \beta_{10}(dmob_{ijt}) + \beta_{11} \ln(dist_{ij}) + \beta_{12}(contig_{ij}) \\ & + \beta_{13}(tdiff_{ij}) + \beta_{14}\left(comlang_{off_{ij}}\right) + \beta_{15}\left(comlang_{ethno_{ij}}\right) \\ & + \beta_{16}(colony_{ij}) \end{aligned}$$

A.3.3. Augmented model with Internet diffusion (eq. 7)

$$\begin{aligned} coinv_{ijt} = {} & \beta_0 + \beta_1 \ln(pat_{it}) + \beta_2 \ln(pat_t) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) \\ & + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8(dint_{ijt}) \\ & + \beta_9 \ln(dist_{ij}) + \beta_{10}(contig_{ij}) + \beta_{11}(tdiff_{ij}) + \beta_{12}\left(comlang_{off_{ij}}\right) \\ & + \beta_{13}\left(comlang_{ethno_{ij}}\right) + \beta_{14}(colony_{ij}) \end{aligned}$$

A.3.4. Augmented model with Average ICT Diffusion (eq. 8)

$$\begin{aligned} coinv_{ijt} = {} & \beta_0 + \beta_1 \ln(pat_{it}) + \beta_2 \ln(pat_t) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) \\ & + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8(avgICT_{ijt}) \\ & + \beta_9 \ln(dist_{ij}) + \beta_{10}(contig_{ij}) + \beta_{11}(tdiff_{ij}) + \beta_{12}\left(comlang_{off_{ij}}\right) \\ & + \beta_{13}\left(comlang_{ethno_{ij}}\right) + \beta_{14}(colony_{ij}) \end{aligned}$$

A.3.5. Augmented model with Average (ln) ICT Diffusion (aux2)

$$\begin{aligned} coinv_{ijt} = {} & \beta_0 + \beta_1 \ln(pat_{it}) + \beta_2 \ln(pat_t) + \beta_3 \ln(pat_{jt}) + \beta_4 \ln(gdptot_{it}) \\ & + \beta_5 \ln(gdptot_{jt}) + \beta_6 \ln(poptot_{it}) + \beta_7 \ln(poptot_{jt}) + \beta_8 ln(avgICT_{ijt}) \\ & + \beta_9 \ln(dist_{ij}) + \beta_{10}(contig_{ij}) + \beta_{11}(tdiff_{ij}) + \beta_{12}\left(comlang_{off_{ij}}\right) \\ & + \beta_{13}\left(comlang_{ethno_{ij}}\right) + \beta_{14}(colony_{ij}) \end{aligned}$$

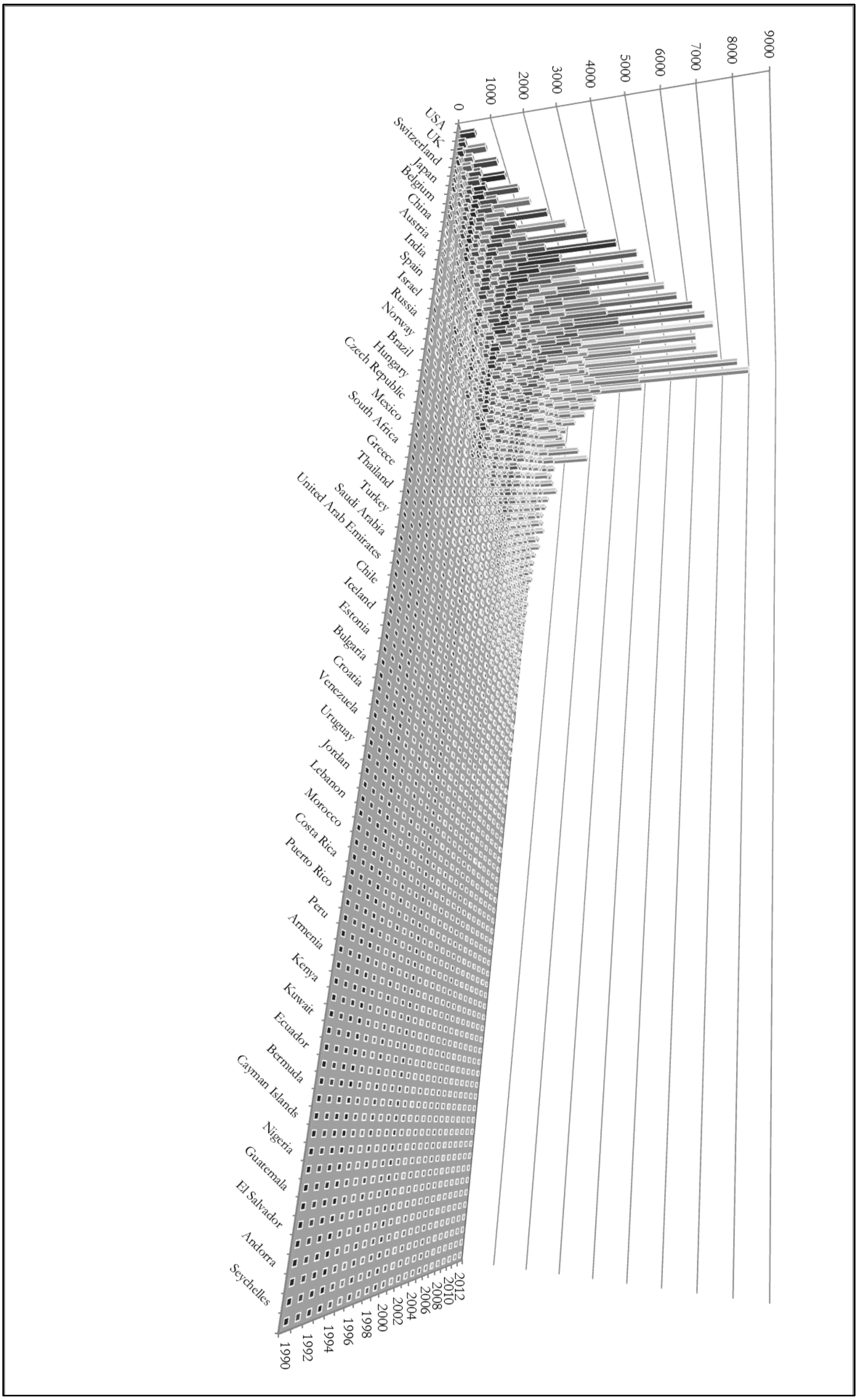
Appendix 3. Final country sample

Australia	Finland	Netherlands
Austria	France	New Zealand
Belarus	Germany	Norway
Belgium	Greece	Poland
Brazil	Hungary	Portugal
Bulgaria	Iceland	Russia
Canada	India	Slovak Republic
Chile	Ireland	Slovenia
China	Israel	South Africa
Colombia	Italy	Spain
Costa Rica	Japan	Sweden
Croatia	Korea	Switzerland
Cyprus	Latvia	Turkey
Czech Republic	Lithuania	Ukraine
Denmark	Luxembourg	United Kingdom
Estonia	Mexico	United States

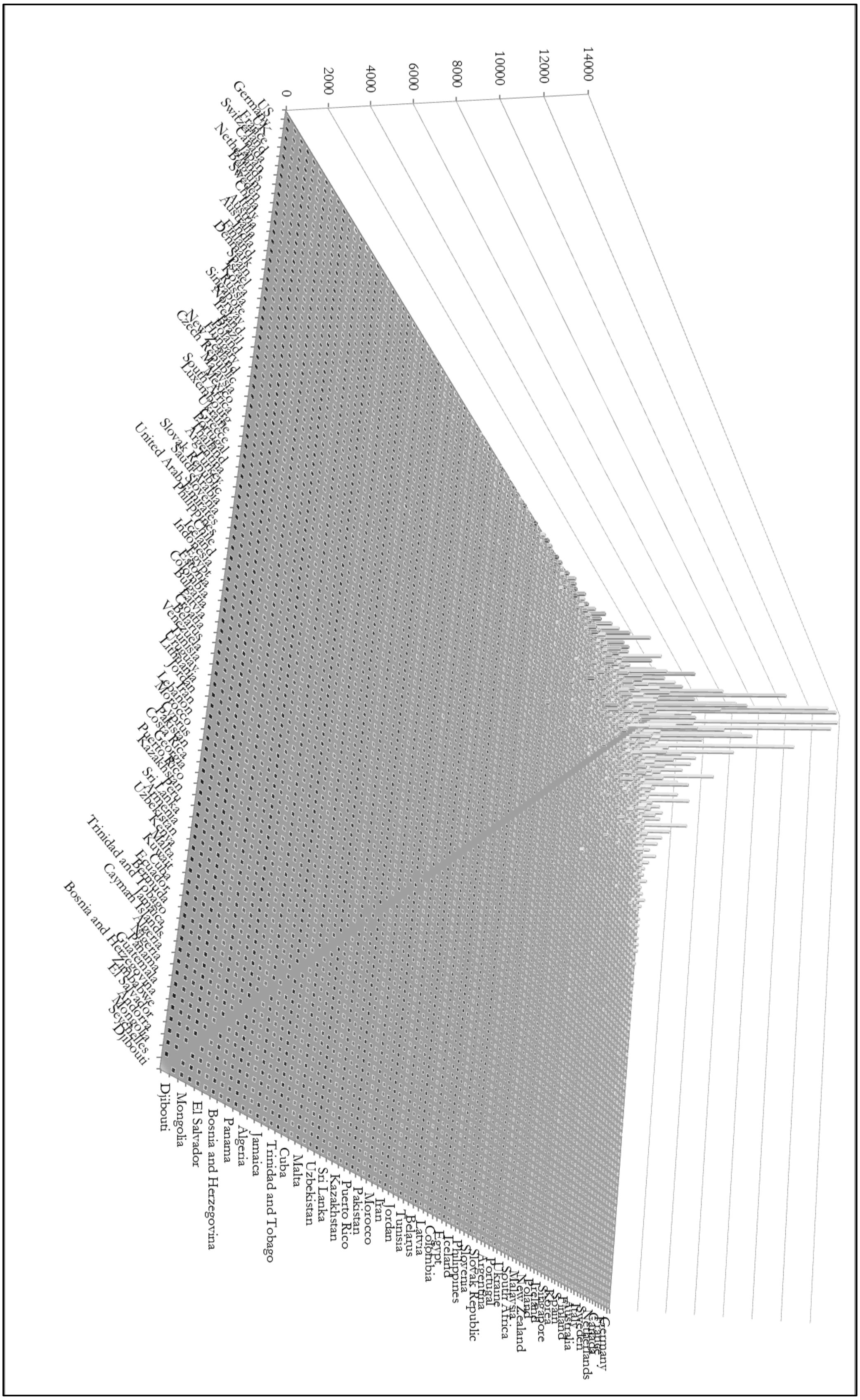
Appendix 4. Spearman correlation matrix: Base levels, adapted from OECD & CEPII

	coinv_ijt	pat_it	pat_jt	gdptot_it	gdptot_jt	poptot_it	poptot_jt	internet_it	internet_jt	telephone_it	telephone_jt	mobile_it	mobile_jt	dist_ij	tdiff_ij	contig_ij	comlang_off	comlang_eth	colony_ij
coinv_ijt	1																		
pat_it	0,5128	1																	
pat_jt	0,5128	0,0373	1																
gdptot_it	0,4372	0,8387	0,0033	1															
gdptot_jt	0,4372	0,0033	0,8387	-0,0115	1														
poptot_it	0,2596	0,5114	-0,0101	0,8216	-0,016	1													
poptot_jt	0,2596	-0,0101	0,5114	-0,016	0,8216	-0,0206	1												
internet_it	0,2685	0,4161	0,1914	0,134	0,0721	-0,1935	0,0083	1											
internet_jt	0,2685	0,1914	0,4161	0,0721	0,134	0,0083	-0,1935	0,6722	1										
telephone_it	0,2925	0,5088	-0,0092	0,2305	-0,0093	-0,2172	0,0017	0,4143	-0,013	1									
telephone_jt	0,2925	-0,0092	0,5088	-0,0093	0,2305	0,0017	-0,2172	-0,013	0,4143	-0,0055	1								
mobile_it	0,1961	0,2975	0,2066	0,0558	0,0822	-0,1646	0,01	0,8769	0,7221	0,227	-0,0189	1							
mobile_jt	0,1961	0,2066	0,2975	0,0822	0,0558	0,01	-0,1646	0,7221	0,8769	-0,0189	0,227	0,7795	1						
dist_ij	-0,1556	0,0006	0,0006	0,1041	0,1041	0,1807	0,1807	-0,0752	-0,0752	-0,1635	-0,1635	-0,1177	-0,1177	1					
tdiff_ij	-0,0389	0,0691	0,0691	0,1708	0,1708	0,2185	0,2185	-0,0421	-0,0421	-0,0877	-0,0877	-0,1097	-0,1097	0,8172	1				
contig_ij	0,1708	0,024	0,024	0,0205	0,0205	0,0247	0,0247	-0,0046	-0,0046	0,0024	0,0024	0,0108	0,0108	-0,3445	-0,242	1			
comlang_off_ij	0,2157	0,0937	0,0937	0,087	0,087	0,0381	0,0381	0,0107	0,0107	0,0661	0,0661	-0,0163	-0,0163	0,0078	0,0041	0,1725	1		
comlang_ethno_ij	0,2155	0,1116	0,1116	0,1064	0,1064	0,0776	0,0776	0,0002	0,0002	0,0504	0,0504	-0,0283	-0,0283	0,0472	0,0279	0,141	0,7952	1	
colony_ij	0,1232	0,0387	0,0387	0,048	0,048	0,0419	0,0419	0,0032	0,0032	0,0217	0,0217	0,0054	0,0054	-0,0998	-0,0358	0,2521	0,305	0,3209	1

Appendix 5. Total co-inventions by country: full sample, adapted from OECD



Appendix 6. Total co-inventions between country pairs: full sample, adapted from OECD

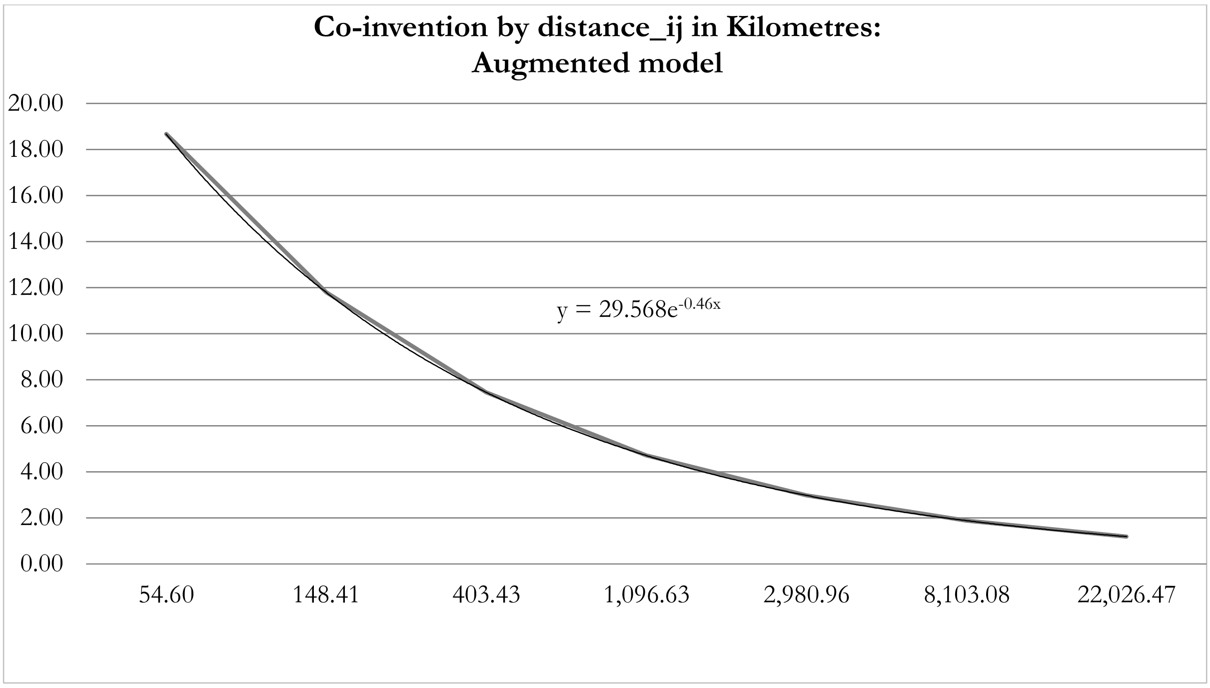


Appendix 7. Results for original sample (92 countries and period 1990-2012)

Explanatory variables	**Eq. 4**	**Aux1**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Aux2**
	Non-Aug	**Aug**	**Aug ICTLvls**	**Aug ICTTrans**	**Aug Int**	**Aug ICTavg**	**Aug LnICTavg**
lnpat_it	0,351***	0,356***	0,362***	0,364***	0,358***	0,362***	0,359***
	(0,0648)	(0,0650)	(0,0700)	(0,0688)	(0,0668)	(0,0684)	(0,0688)
lnpat_jt	0,351***	0,356***	0,362***	0,364***	0,358***	0,362***	0,359***
	(0,0648)	(0,0650)	(0,0700)	(0,0688)	(0,0668)	(0,0684)	(0,0688)
lngdptot_it	0,696***	0,688***	0,658***	0,650***	0,679***	0,665***	0,673***
	(0,209)	(0,210)	(0,204)	(0,227)	(0,213)	(0,221)	(0,221)
lngdptot_jt	0,696***	0,688***	0,658***	0,650***	0,679***	0,665***	0,673***
	(0,209)	(0,210)	(0,204)	(0,227)	(0,213)	(0,221)	(0,221)
lnpoptot_it	-0,305	-0,196	-0,0834	-0,107	-0,183	-0,177	-0,188
	(0,257)	(0,254)	(0,264)	(0,258)	(0,259)	(0,260)	(0,263)
lnpoptot_jt	-0,305	-0,196	-0,0834	-0,107	-0,183	-0,177	-0,188
	(0,257)	(0,254)	(0,264)	(0,258)	(0,259)	(0,260)	(0,263)
internet_it			0,0378				
			(0,0922)				
internet_jt			0,0378				
			(0,0922)				
telephone_it			-0,0027				
			(0,186)				
telephone_jt			-0,0027				
			(0,186)				
mobile_it			0,0733				
			(0,07)				
mobile_jt			0,000733				
			(0,07)				
dint_ijt				0,103	0,104		
				(0,0982)	(0,102)		
dtel_ijt				0,0817			
				(0,155)			
dmob_ijt				-0,253***			
				(0,0831)			
avgdICT_ijt						-0,149	
						(0,183)	
lnavgdICT_ijt							-0,0343
							(0,0346)
lndist_ij	-0,460***	-0,203***	-0,195***	-0,181***	-0,194***	-0,192***	-0,190***
	(0,0297)	(0,0541)	(0,0535)	(0,0540)	(0,0537)	(0,0535)	(0,0526)
tdiff_ij		-0,0153	-0,0181	-0,0171	-0,0184	-0,0176	-0,0176
		(0,0132)	(0,0131)	(0,0130)	(0,0131)	(0,0131)	(0,0132)
contig_ij		0,411***	0,410***	0,417***	0,416***	0,405***	0,404***
		(0,0786)	(0,0781)	(0,0786)	(0,0783)	(0,0784)	(0,0793)
comlang_off_ij		0,646***	0,657***	0,665***	0,658***	0,658***	0,658***
		(0,103)	(0,102)	(0,0999)	(0,102)	(0,101)	(0,101)
comlang_ethno_ij		0,0214	0,0153	0,0190	0,0104	0,0198	0,0215
		(0,101)	(0,1000)	(0,0989)	(0,101)	(0,100)	(0,101)
colony_ij		0,00682	-0,00445	-0,0187	-0,00431	-0,00742	-0,00819
		(0,0573)	(0,0569)	(0,0561)	(0,0567)	(0,0563)	(0,0561)
Constant	-14,18***	-15,33***	-14,55***	-14,78***	-15,24***	-15,13***	-15,38***
	(1,582)	(1,594)	(1,468)	(1,669)	(1,606)	(1,671)	(1,599)
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	149 974	149 974	137 648	137 648	137 648	137 648	137 648
R-squared	0,900	0,953	0,953	0,955	0,953	0,954	0,954

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Numbers based on model specification
Effects for countries *i* and *j*

Appendix 8. Margins Co-invention (model 4), adapted from OECD & CEPII, original sample



Appendix 9. Margins Co-invention (model aux1), adapted from OECD & CEPII, original sample