WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

EFMD
EQUIS
ACCREDITED

Bachelor Thesis

Analysing the Impact of Corruption on Innovation Performance in European Regions

Jeremias Braendle

Date of Birth: 02.10.1999
Student ID: h11817039

Subject Area: Information Business

Studienkennzahl: J123456789

Supervisor: Johannes Wachs, PhD

Date of Submission: 30 May 2021

Department of Information Systems and Operations, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

DEPARTMENT FÜR INFORMATIONS-VERARBEITUNG UND PROZESS-MANAGEMENT DEPARTMENT OF INFORMATION SYSTEMS AND OPERATIONS

Contents

List of Figures

List of Tables

Abstract

The aim of this thesis is to analyse the relationship between corruption and innovation performance in European regions. It differentiates itself from and adds to the existing literature by considering not only the quantity but also the respective quality of patents. In addition, the impact of corruption on technological complexity and the technological composition of regions is investigated. To improve the validity of the results, two different corruption proxies are used, namely the rate of single bidding in open tenders and the corruption pillar of the European Quality of Government Index (EQI). Regarding the methodology, both the between and within estimators are applied to isolate the cross-sectional and time dimension of variation in the data. The results provide additional evidence supporting the argument in the literature that corruption sands the wheels of innovation. This is true not only for the quantity of patent applications, but also for their quality and the technological complexity of regions. Some evidence was found that corruption might also alter the technological composition of regions, leading to a higher share of simple technologies. The findings imply that the current decline in the control of corruption in Europe must be halted to clear the way for bringing Europe back to the forefront of innovation.

1 Introduction

Corruption, being defined as the abuse of entrusted power for one's own personal gain, has existed for thousands of years and is a global phenomenon that occurs in all cultures and societies. Its thorough exploration in research, on the other hand, has begun only about 25 years ago. Nevertheless, the economic and societal implications of corruption are quite well researched [60]. Besides the monetary costs incurred by corruption, Transparency International also defines political costs, since the freedom and rule of law are affected, social costs because corruption decreases the trust and participation in government and environmental costs because corruption facilitates environmentally harmful business practices [34].

A specific relationship that has been examined by several researchers already is that between corruption and innovation performance, both at the (sub-) national and firm-level. Interestingly enough though, there is no generally accepted truth regarding the polarity of this relationship. Consequently, there are two different views in the literature, often called the "greasing the wheels of innovation" and "sanding the wheels of innovation" view [48]. Section 2 will elaborate on this discourse in more detail.

A less debated fact is that innovation has a positive effect on economic growth [29, 50, 26]. The European Commission has recognized innovation not just as one of the key drivers for economic growth and progress but also as a means to "deliver jobs, prosperity, quality of life and global public goods" [17]. Consequently, the programme "Horizon 2020" was initiated to secure Europe's global competitiveness. It was active from 2014 to 2020 with a budget of almost EUR 80 billion. The goal of this research and innovation centred programme is described by the European Commission as follows.

> "The goal is to ensure Europe produces world-class science and technology, removes barriers to innovation and makes it easier for the public and private sectors to work together in delivering solutions to big challenges facing our society." [18]

The research question if and to what extent corruption affects the innovation performance of European regions has relevant implications for policy-making. If the hypothesis that corruption is a major obstacle to innovation proves true, anti-corruption policy initiatives could be an effective measure to support the long-term development of European regions and come closer to the goal set by the European Commission.

Although this research question has been dealt with before, there are still research gaps that justify its re-examination [52]. On the data side, there has been some progress in the objective measurement of corruption risk in the form of red flag indicators based on public procurement data, whose main advantage is that they are available for all NUTS 2 regions of the EU for several years and that they are not based on perceptions but on objective, measurable facts [24]. Moreover, while there is good theoretical and empirical evidence on the relationship between corruption and the number of patent applications, it has not yet been analysed whether this relationship extends to other facets of innovation performance such as the quality of patents, the technological complexity, and the technological composition of regions.

2 Literature Review

In this section, the background literature will be reviewed. The first subsection deals with the question, how corruption can be reliably measured. The second subsection will cover a similar question but for the innovation performance of regions. The final subsection discusses the literature on the relationship between corruption and innovation.

2.1 Measuring Corruption Risk

Measuring corruption is a challenging task, especially when it comes to grand corruption, that is, the dealings of the political elite and businesses. While petty corruption, describing corruption in the interaction between public officials and ordinary citizens, can be measured rather easily by surveying citizens who have potentially experienced or witnessed this form of corruption before, it is significantly more difficult to detect grand corruption like favouritism in public procurement [41].

2.1.1 Perception Based Corruption Indicators

Two of the most established attempts at measuring corruption are the World Bank's Control of Corruption indicator (CC) and the Corruption Perceptions Index (CPI), which is calculated each year by Transparency International [43]. Both are composite indices based on expert's and business executive's perception of corruption in the respective country. While these perception-based indices have opened the door for more elaborate empirical research on corruption, their validity is somewhat limited, and they do not allow for fine-grained analyses since they are only available at the country level [40].

Charron et al. (2014) addressed the lack of a reliable corruption indicator on a sub-national level by conducting a multi-country survey questioning citizens from 172 European regions about their perceptions of their quality of government. From the results, they derived the European Quality of Government Index (EQI), a composite index describing the level of "low corruption, impartial public services and the rule of law". They found that there are significant disparities in the quality of government within European countries. The Italian region Bolzano for example, ranks among the very best in Europe, while Campania is one of the regions with the lowest quality of government in Europe. The difference between the two regions is bigger than, for example, the difference between Denmark and Hungary [14]. The EQI was constructed four times so far, building on data collected in 2010, 2013, 2017, and 2021 [16]. While the EQI definitely contributes to the empirical research on regional corruption, it still suffers from the problems of perception-based indices and is only available for four non-subsequent years, which impedes analyses over time.

2.1.2 Objective Corruption Risk Indicator

Until recently, attempts at creating objective measures of corruption have lacked validity and suffered from low data availability [40]. In 2016, Fazekas et al. proposed a new composite indicator of institutionalized grand corruption. Their Corruption Risk Index (CRI) addresses the lack of valid and widely available objective corruption indicators. The CRI is based on measuring the presence of so-called "red flags" in open tenders. These red flags can hint at favouritism and the bypassing of fair competition in all phases of the tendering process. In the submission phase, the set of bidders can be intentionally restricted. In the assessment phase, bidders can be assessed unfairly, and in the delivery phase, the conditions of performance can be modified ex-post. The authors use logistic and linear regressions as validation methods, linking potential corruption inputs, which are defined as techniques to achieve the corrupt outcomes, to the respective likely corruption outcomes while controlling for influencing factors like the general competitiveness of the market [24]. Decarolis and Giorgiantonio (2020) have further contributed to the validation of red flags as predictors of corruption in public procurement through the use of machine learning models. Their primary measure of direct corruption risk is based on police investigations regarding corruption-related crimes of firm's executives. They conclude that many of the red flag indicators are valid predictors of corruption and propose a set of additional red flags [21].

2.2 Measuring Innovation Performance

Measuring the innovation performance of regions has been done frequently in the literature. However, it is often unclear why a certain measure was chosen and whether it is a valid proxy for the research question at hand. Brenner and Broekel (2009) argue that the characteristics of a region affect its innovation performance in two ways. First, it can attract innovation generators and second, it can facilitate the innovation process of existing innovation generators. One of the most common approaches to measuring the innovation performance of a region is to measure the total innovation outcome, usually approximated by the number of patent applications. In this approach, the two aforementioned capacities are not measured separately. Instead, the region with all of its characteristics is treated as a black box [9].

2.2.1 European Innovation Scoreboard

An alternative approach to measuring the innovation performance of European countries and regions was introduced in 2001 by the European Commission in the form of the European Innovation Scoreboard (EIS) and the Regional Innovation Scoreboard (RIS). The 2020 version of this composite index comprises 27 indicators, capturing the framework conditions, investments, innovation activities and impacts. The main goal of the scoreboard is to help countries improve their innovation performance by demonstrating which areas they should focus their efforts on [32].

The methodology of the EIS has been revised several times over the years as a response to criticism and changed circumstances [30, 31]. While it is laudable that the European Commission acknowledges possible improvements, it means that the EIS is not really comparable over time. Furthermore, the general approach has been criticised because it combines innovation inputs and outcomes in one composite indicator [22]. Brenner and Broekel (2009) argue that this approach makes it unclear what the EIS and RIS really measure. They conclude that, although it is not explicitly stated, the RIS might be an attempt at capturing a region's capacity to establish innovation generators by including many different proxies [9].

2.2.2 Patents as Measure of Innovation Outcome

Even if innovation inputs and outcomes are kept separately, the commonly available proxies have some substantial limitations. For example, the common practice of using the number of patents as a proxy for innovation outcome is quite imperfect. The propensity to patent differs between industries, and many inventions are not patented at all, especially process innovations. The unique advantage, however, is that patent data is widely available at the regional level over many years [52].

One might argue that measuring innovation performance simply through the number of patent applications postulates that all patents are the same. This, however, does not reflect reality since every patent has its unique economic and technological value. To address this fact, scholars have proposed several methods to evaluate patents. Van Zeebroeck and van Pottelsberghe (2011) criticize the robustness and inconsistencies between studies of many of these proposed measures, concluding that great care should be taken when using a single patent value indicator [56]. Similarly, van Zeebroeck (2007) proposes that composite indicators can better capture the different dimensions of a patent's value than single indicators since the respective single indicators are only weakly correlated with each other and show different industrial patterns [55]. An important contribution to this strain of research has been made by Squicciarini et al. (2013). They propose a wide array of patent value indicators and accompany them with sensitivity and correlation tests and, most notably, a dataset that contains these indicators for patent applications filed at the European Patent Office (EPO) [53].

2.2.3 Technological Complexity

Mewes and Broekel (2020) have shown that the technological complexity of a region is a potent predictor of economic growth. Complex technologies are difficult to invent and imitate but offer substantial economic benefits [45]. It can thus be argued that the capacity of a region to attract generators of complex technological innovations is a good measure for its innovation performance. The potential problem with this argument is that the agglomeration of industries can not be seen as a direct reflection of a region's specific attractiveness but is strongly influenced by chance, self-reinforcing dynamics and historical developments [9].

2.3 Relationship Between Corruption and Innovation Performance

As briefly touched upon in the introduction, the potential effects of corruption on innovative capacity have been investigated already by several research scholars. Interestingly, there are two opposing views regarding the polarity of this relationship.

2.3.1 Greasing the Wheels of Innovation

One line of argumentation is that corruption can be beneficial for economic growth and innovative activity because it reduces the amount of time and effort spent on bureaucratic processes for companies that are willing to pay bribes. This "grease money" effect has, for example, been studied by Karaman (2018). She analysed the effect of corruption on firm-level innovation in Eastern European and Central Asian countries and concluded that there is a significant and robust positive relationship [36]. Riaz and Cantner (2019) have analysed 16 developing and emerging economies and found that in the majority of cases, there is a positive association between monetary corruption and innovative performance of firms, although the effect varies by industry [51].

2.3.2 Sanding the Wheels of Innovation

The predominant view, however, is that corruption has a negative impact on innovation performance. Corrupt regimes are unattractive for outsiders because they cannot profit from the established network structure and face unfair conditions. Furthermore, the rent-extraction of corrupt officials reduces the prospected profits of potential innovators. Tampubolon (2018) adds that the already high risk involved in investing in research and development gets amplified by a low quality of governance because the risk of successful arbitrary claims to the intellectual property against the inventors' claims is higher if no fair and reliable institution upholds the law [54].

Rodríguez-Pose and Di Cataldo (2015) have analysed how the quality of governance affects innovation outcome in European regions and concluded that there is a significant positive relationship, or in other words, the lower the quality of government, the less innovative a region is. They employed a dynamic fixed effects regression model using the annual change in the logarithmic transformation of the number of patent applications to the European Patent Office per million inhabitants as the dependent variable and the EQI as the independent variable while controlling for time-related shocks, R&D expenditure, the characteristics of the labour market, and the socio-economic structure of the regions [52].

One aspect of this study that needs to be critically highlighted is that Rodríguez-Pose and Di Cataldo chose a time span between 1997 and 2009 for their analysis even though at that point in time, the EQI was only available for the year 2010. Referencing the work of Charron et al. (2014), they assumed that within countries, the change in institutional quality is homogeneous enough to take the variation of the World Governance Index (WGI), which is calculated at the country level, to extend the EQI across longer time-periods [52]. The question arises whether this approach does not undermine the original intention of acknowledging the substantial within-country variations in the quality of government. Regarding the research question, research subjects, and general methodology, this work is closest to my thesis.

The next section explains the variables used in this study, which will also shed more light on how this thesis differs from previous research.

3 Data

3.1 Dependent Variables - Proxies for Innovation Performance

As was elaborated in section 2, there is no real state of the art when it comes to measuring innovation performance. I do share the criticism of Brenner and Broekel (2009) regarding the EIS. Not keeping innovation outputs and inputs as separate measures is imprecise and impedes the interpretation of findings. That said, only focusing on the innovation output might also fall short.

Broekel et al. (2017) argue that besides the total innovation output, an important part of innovation performance is how many resources were utilized to achieve this result. Their contribution to the literature on this matter is that they developed a measure for the cross-industry innovation efficiency of regions while controlling for the respective industrial structure [12].

Regarding the underlying research question of this thesis, both the innovation effectiveness and the innovation efficiency are of interest. The latter describes how well the given innovation inputs are transformed into innovation outputs. Therefore, this measure is suitable to answer whether corruption facilitates or impedes the innovation process of existing innovation generators. The total innovation output, on the other hand, is more suitable if we also want to assess a region's capacity to attract innovation generators in the first place. To allow for a comprehensive understanding, both concepts are modelled. My source of patent applications to the European Patent Office (EPO) is the OECD REGPAT Database, January 2021.

3.1.1 Patent Quality

In addition to the number of patent applications, I also consider their respective quality, as the considerable differences in the economic and technological relevance of different patents should not be ignored. The patent quality proxy used in this thesis was developed by Squicciarini et al. (2013). I use their composite index, which is based on the following four components.

i Number of forward citations
Forward citations are citations that a patent receives. They are, therefore, a measure of the technological importance because subsequent technologies build upon the cited patent. For the composite indicator, forward citations are counted over a five year period after the patent was published, which is usually 18 months after the filing date. This implies a timeliness problem due to truncation. Time-fixed effects should, however, be able to capture and compensate for this effect. Empirically, a rather small fraction of patents receive a large proportion of all forward citations.

ii Patent family size
The patent family size is defined as the number of different patent offices to which an invention has been filed. This is a good indicator for the economic value of a patent because applicants will only accept these additional costs and the extra time involved if they assume that the profit generated by the patented technology will exceed the investment

sum. The exceptionally high value associated with large patent families has been shown by Harhoff et al. (2003) [20].

iii Number of claims
Claims define which aspects of a technology are legally protected by the patent. A higher number of claims usually imposes a higher application fee but is also associated with a higher economic value [42].

iv Patent generality index
The patent generality index is also based on forward citations. It aims at measuring the relevance of a patent for later inventions in different technological fields. If a patent only receives forward citations in its own technological field, the generality is considered to be low. A limitation of this index is that it does not differentiate according to the distance of technology fields. Similar technologies are treated in the same way as very distant ones.

The patent quality index weights all four components equally because the empirical evaluation has shown that unequal weights would have to differ between technologies and time frames which would limit the comparability of the indicator [53]. The data are included in the OECD Patent Quality Indicators Database.

3.1.2 Technological Complexity

One of my research questions is whether and to what extend corruption affects technologies of different complexity levels differently. But how can technological complexity be measured and quantified? Broekel (2019) tries to answer this question using structural diversity. He models technologies as combinatorial networks and measures the diversity of the occurring topologies. To demonstrate the reasoning behind this approach, let's think of a chair as an example of a very simple technology. A typical chair consists of a seat, four chair legs and a backrest. It can be modelled as a star-like structure with the seat at the centre. Consequently, only minimal information is required to describe the chair from an information-theoretical perspective, and it can be easily invented, copied, and codified. In contrast, it is far more difficult to describe the network topology of a rocket engine, for example. Broekel (2019) therefore argues that the structural diversity of a technology can be used as a proxy for its complexity [10]. Broekel (2019) calculated the structural diversity score for 655 technologies, defined by the four-digit CPC classes. The data containing the respective values for each of the 655 technology classes for the years 1970 to 2016 are provided on his website [11].

3.2 Proxies for Corruption

As was explained in section 2, there is no ideal measure for corruption. When it comes to the regional level, the number of eligible indicators shrinks even more.

3.2.1 European Quality of Government Index (EQI)

Besides suffering from the problems associated with perception-based indicators and only being available for four non-subsequent years, the EQI is arguably the most sophisticated attempt to date to measure corruption risk in European regions. Since it is not available for subsequent years, its applicability in panel analyses is unfortunately somewhat limited. However, since corruption is a rather persistent phenomenon and big jumps between successive years are not expected, interpolating the values for the missing years is a reasonable approach [28].

3.2.2 Red Flags in Public Procurement

In addition to the EQI, the CRI and other variants of red flag based corruption risk indicators were considered as corruption proxies.

To cross-check the data, I used two different sources. First, I drew on opentender.eu, an online platform administered by DIGIWHIST aiming to make public procurement more transparent [1]. DIGIWHIST is a project carried out by six European research institutes and funded through Horizon 2020 to collect, structure, analyse, and disseminate public procurement data [44]. To be precise, I use their collection of contracts published on the Tenders Electronic Daily (TED) database. It is mandatory to publish tenders on TED if they fall within the scope of the EU Public Procurement Directives, which applies to tenders whose contract value exceeds around EUR 130,000 if they are service contracts and around EUR 5,000,000 if they are public works contracts. DIGIWHIST also collect national public procurement data under the threshold value, but they are generally not comparable due to the different national tendering procedures [23].

The second source for the corruption proxies is the QOG EU Regional Dataset, curated by the Quality of Government Institute at the University of Gothenburg and contains more than 300 variables covering twelve categories. It contains the EQI and its components as well as four of the proxies for corruption risk in public procurement, namely the share of single bidding in competitive markets, the share of contracts with no published call for tender, the share of contracts with a procedure classified as non-open, and the share of contracts with tax haven red flag [13].

Due to some data inconsistencies in the dataset containing individual contracts, the QOG EU Regional Dataset, which is aggregated on the regional level, was used in the final analysis. Unfortunately, the earliest year covered in the data is 2011, limiting the temporal dimension of the analysis to a period of less than ten years, which admittedly is not very much in terms of corruption, which generally tends to be quite persistent. Nevertheless, there are observable trends, even over the course of just a few years [15].

3.3 Control Variables

To minimize omitted variables bias, I control for the main potential confounders. Based on the literature, I identified the following variables that define the socioeconomic and labour market structure of regions and the two main innovation inputs, namely human capital, represented by the share of employees in high-tech sectors, and monetary capital, i.e. spending on research and development.

Table 1: Control Variables

	Definition	**Source**
GDP PPS	Gross domestic product in purchasing power standards per inhabitant	Eurostat
Unemployment	Unemployed persons divided by economically active population * 100	Eurostat
Pop. Density	Inhabitants per square kilometre	Eurostat
Primary Sector	Workers in the primary sector, divided by total number of workers * 100	Eurostat
Tertiary Ed.	Persons with tertiary education aged 25–64 divided by total population aged 25–64	Eurostat
Hightech	Employees in high-technology manufacturing and knowledge-intensive high-technology services, divided by total number of workers * 100	Eurostat
R&D Spending	Intramural R&D expenditure (GERD) in Euro per inhabitant	Eurostat

4 Data Processing

4.1 NUTS Nomenclature

The Nomenclature of Territorial Units for Statistics (NUTS) was first established by Eurostat and used in EU legislation in 1988 to facilitate consistent statistical analyses and political interventions on a regional level. 15 years later, in 2003, the European Parliament and the Council formally encoded it in a regulation. Since then, there have been multiple amendments, namely in 2006, 2010, 2013, 2016, and most recently, in January 2021. Regions are classified in a three-level hierarchy, although for some countries, these levels represent the same territory. In the case of Luxembourg, for example, there is no difference between NUTS 1, NUTS 2 and NUTS 3, which is reflected in the respective codes LU0, LU00 and LU000. Table 2 depicts the Austrian "Bundesland" Vorarlberg as a typical example of the hierarchical classification [2].

Table 2: NUTS Classification of Vorarlberg - Austria

	NUTS1	**NUTS2**	**NUTS3**
AT3	Westösterreich		
AT34		Vorarlberg	
AT341			Bludenz-Bregenzer Wald
AT342			Rheintal-Bodenseegebiet

Whenever possible, NUTS regions are based on administrative units, like the Austrian "Bundesländer". However, sometimes there is no administrative unit for a particular level, so the aggregation of smaller administrative units is necessary. This is, of course, not ideal for statistical analyses, but due to limited data availability, this constraint has to be accepted.

The NUTS' main purpose is to ensure a consistent, harmonized standard for the classification of regions. However, exactly this comparability is limited across several years and NUTS versions. Amendments include recoding, discontinuation, merging, splitting, and boundary changes of regions [2]. Therefore, harmonising different NUTS versions is challenging, especially if more than two versions are included in the data.

To handle this issue, the two R packages *Eurostat* and *regions* were used. The *Eurostat* package implements several tools and functions specifically designed to work with data from Eurostat. Some of these functions are dedicated to harmonizing NUTS 2013 and NUTS 2016 codes [39]. Recoding and relabelling between these two versions are handled well by the package. However, harmonizing regions that have undergone spatial changes and dealing with NUTS versions before NUTS 2013 is not possible. The *regions* package was developed specifically to address the typical data processing, validation and imputation issues that arise when working with sub-national data. It has extended recoding capabilities compared to the *Eurostat* package, tracking boundary changes in the EU between 1999 and 2021. Furthermore, it allows for data aggregation and disaggregation to impute higher NUTS levels to lower ones and vice versa [6].

Unfortunately, even with both packages employed in an iterative process, the harmonization proved cumbersome and error-prone. Therefore, some regions had to be excluded from the analysis to ensure consistent units for which a comparison over time is meaningful. Furthermore, choosing the NUTS version for the harmonization was not straightforward because while the OECD REGPAT database mainly corresponds to NUTS 2013, other used data sources mainly follow NUTS 2016. NUTS 2013 was finally chosen due to the importance of the REGPAT data for the analysis.

4.2 Patent Data

The OECD REGPAT database, January 2021, contains several datasets derived from the European Patent Office's Worldwide Statistical Patent Database (PATSTAT Global, Autumn 2020). Two of these datasets are used in this thesis. The first contains patent applications and inventors, as well as their residential addresses and, derived from the addresses, the respective NUTS 3 regions. It consists of 9,768,237 observations. The second relevant dataset contains patent applications and the Corporate Patent Classification (CPC) classes related to the invention. CPC classes define technologies hierarchically and classify them into nine classes on the highest and more than 230,300 subclasses at the most granular level. The CPC classes are needed to link the technological complexity to the patents [10]. As patents usually comprise several different CPC classes, this dataset contains significantly more entries than the basic dataset, namely 49,635,779. The third OECD patent dataset used is the OECD Patent Quality Indicators database, January 2021, which contains 19 indicators for 1,449,688 patents.

In the first processing step, the 10-digit CPC classes are reduced to 4-digit CPC classes to align with the work of Broekel (2019). Then, the CPC codes are merged with the patent quality dataset and the respective structural complexity values. Finally, the most complex technology class related to the invention is taken as the complexity score of the patent. This makes more sense than taking the mean value, as otherwise, the complexity of an invention involving complex and simple technologies would be underestimated [45].

After these basic operations, the technology field is determined according to the WIPO IPC - Technology Concordance Table, which divides technologies into 5 sectors and 35 fields. The coarsest level of separation consists of the sectors *Electrical Engineering, Instruments, Chemistry, Mechanical Engineering* and *Other Fields.*

The data is then aggregated to region-year units by summing up the number of patent applications per sector in each year and region weighted by the inventors' share. The patents are linked to a certain region, not via the applicant but the inventors. This is more adequate because large corporations often file patent applications through their headquarter and not the subsidiary where the technology was actually invented [45]. When multiple inventors from different regions are involved, the patent is partially attributed to each region. This means that if there is one inventor from region A and one inventor from region B, the patent will only be counted as 0.5 patents for each region to avoid double-counting. After the aggregation, the NUTS 3 codes were harmonized as diligently as possible and aggregated to the NUTS 2 level.

Incompleteness is an issue that runs through all the data used in this thesis. Before aggregation, about one-third of the patent applications did not have a patent quality score assigned to them. After the aggregation, this was reduced to 65 out of 1840 region-year units that fully lacked patents with quality score. The missing values were, if possible, imputed by inserting the mean value of the respective region over all years. This was possible for all but the two regions Vóreio Aigaío and Madeira. The country-level mean was imputed in these cases.

One of the main purposes of the NUTS classification is to have comparable territory units. Nevertheless, there are enormous differences regarding population, area, economic weight and administrative power. The NUTS 2 region with the largest number of inhabitants is Île de France, with about 12.2 million inhabitants, while the smallest region, Åland (Finland), only has about 29,500 inhabitants [2]. This is a ratio of about 414:1. It is therefore indispensable to transform the number of patent applications to a per capita measure.

Unfortunately, the Eurostat population data is systematically missing observations due to the NUTS amendments. To compensate for this, the data was supplemented with data from the Annual Regional Database of the European Commission's Directorate General for Regional and Urban

Policy (ARDECO) which is mainly based on data from Eurostat but updated twice a year and complements the Eurostat data with other national and international sources to fill in missing values [27].

4.3 Corruption Risk Data

Since the EQI is only available at NUTS 1 instead of NUTS 2 level for some regions, the first data processing step replaces the missing values on the NUTS 2 level with the corresponding NUTS 1 level data. In the next step, missing values of the EQI and Red Flag variables are, if possible, linearly interpolated and end values carried backwards or forward, respectively. At this point, there are still missing values because the interpolation is only possible if there is at least one non-missing observation. If this is not the case, the respective NUTS 1 value or country-level value is imputed.

Table 3: Missing Values - Corruption Variables

Variable	**Missing Values in Percent**			
	Initial	Interpol.	NUTS 1	NUTS 0
No Call	22	3.64	1.82	0.36
Non Open	22	3.64	1.82	0.36
Single Bidding	22	4.00	1.82	0.36
Tax Haven	22	3.64	1.82	0.36
EQI Corruption	75	7.64	6.18	2.91

4.4 Control Variables

Eurostat is the main source for the control variables. Eurostat data is structured but not completely tidy. Therefore, a wrangling function was written to process the different datasets efficiently. The function first harmonizes the NUTS codes, then filters for NUTS 2 regions, removes inconsistent regions, selects only the relevant columns containing the region code, year and value of interest, replaces *:* with *NA*, removes whitespace from numbers, converts them to numeric, renames the *Value* column to a meaningful name and removes duplicates.

Missing values are treated in an iterative process. First, if possible, missing values are linearly interpolated. End values are carried forward or backwards, respectively. This interpolation method cannot impute all missing values because for some regions and variables, all values are missing across all years. In these cases, the value of the respective NUTS 1 region was imputed and, if this also failed, the value of the respective country was taken instead. Table 4 shows the percentage of missing values at the beginning and after interpolation and NUTS 1 imputation. After the final step, i.e. the imputation of country-level values, all missing values have been imputed. Table 4 shows the percentage of missing values after each imputation step.

Table 4: Missing Values - Control Variables

Variable	**Missing Values in Percent**		
	Initial	Interpol.	NUTS 1
GDP PPS	16	1.67	1.32
Unemployment	11	0.36	0.36
Pop. Density	10	0.00	0.00
Primary Sector	16	3.28	1.46
Tertiary Ed.	11	0.36	0.00
Hightech	18	4.74	1.09
R&D Spending	36	2.19	1.82

5 Descriptive Analysis

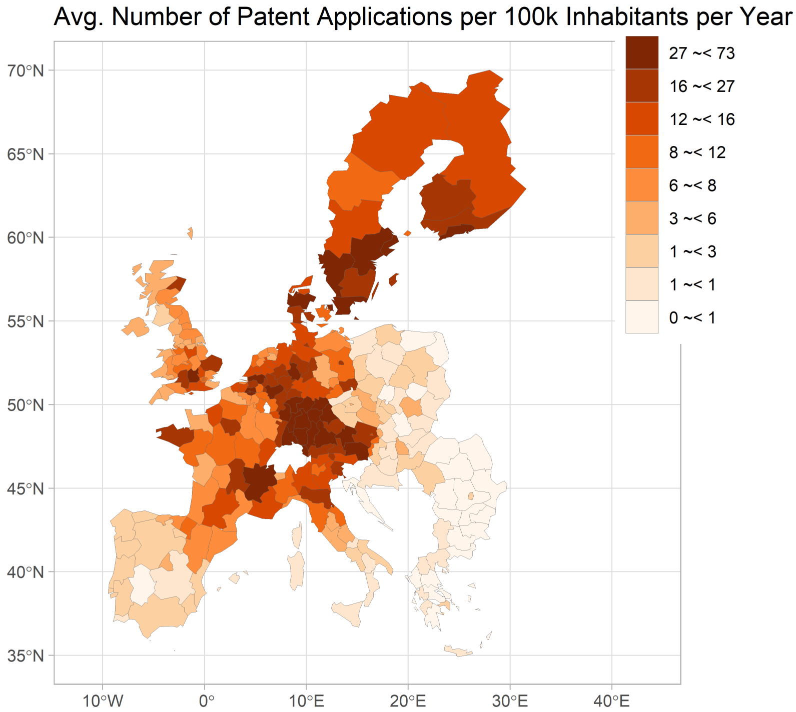
5.1 Innovation Performance

Innovation performance is very heterogeneous in the European Union. As figure 1 shows, patenting is especially concentrated in the southern part of Germany, Austria, Paris, London and parts of the Scandinavian countries. The periphery of Europe, on the other hand, has significantly fewer patent applications per capita. Due to these vast differences, the map's colour coding is based on quantiles instead of equal intervals.

The average patent quality, on the other hand, depicts a slightly different picture, as can be seen in figure 2. The country that stands out due to an exceptionally high average patent quality throughout all its regions is Great Britain. The European periphery tends to produce lower quality patents, although there are some exceptions.

From a methodological viewpoint, it can be argued that the average patent quality might not be the best measure of a region's capacity to produce high-quality patents. If a region produces the highest-quality patents but also simpler ones, the average underestimates its overall innovative capabilities [45]. Figure 2 highlights that the highest-quality patents are indeed

Figure 1: Map of Patent Applications



mainly produced in the centre of Europe compared to the periphery with a significantly lower capacity to produce highly valuable patents.

Looking at the technological complexity, as defined by Broekel (2019), it becomes even clearer how vastly different results and conclusions can be, depending on whether the average or the maximum is considered. Interestingly, the average technological complexity is generally higher in the periphery than in the centre of Europe. However, the most complex technologies are invented mostly in the more central European regions and the Scandinavian countries.

The difference between the average and highest patent quality and complexity is also reflected figure 4, the correlation matrix of the innovation proxies. The average patent complexity is barely correlated with the other patent variables. The maximum patent quality and complexity, on the other hand, are highly correlated with each other and the number of patent applications per 100,000 inhabitants.

Figure 2: Maps of Patent Quality

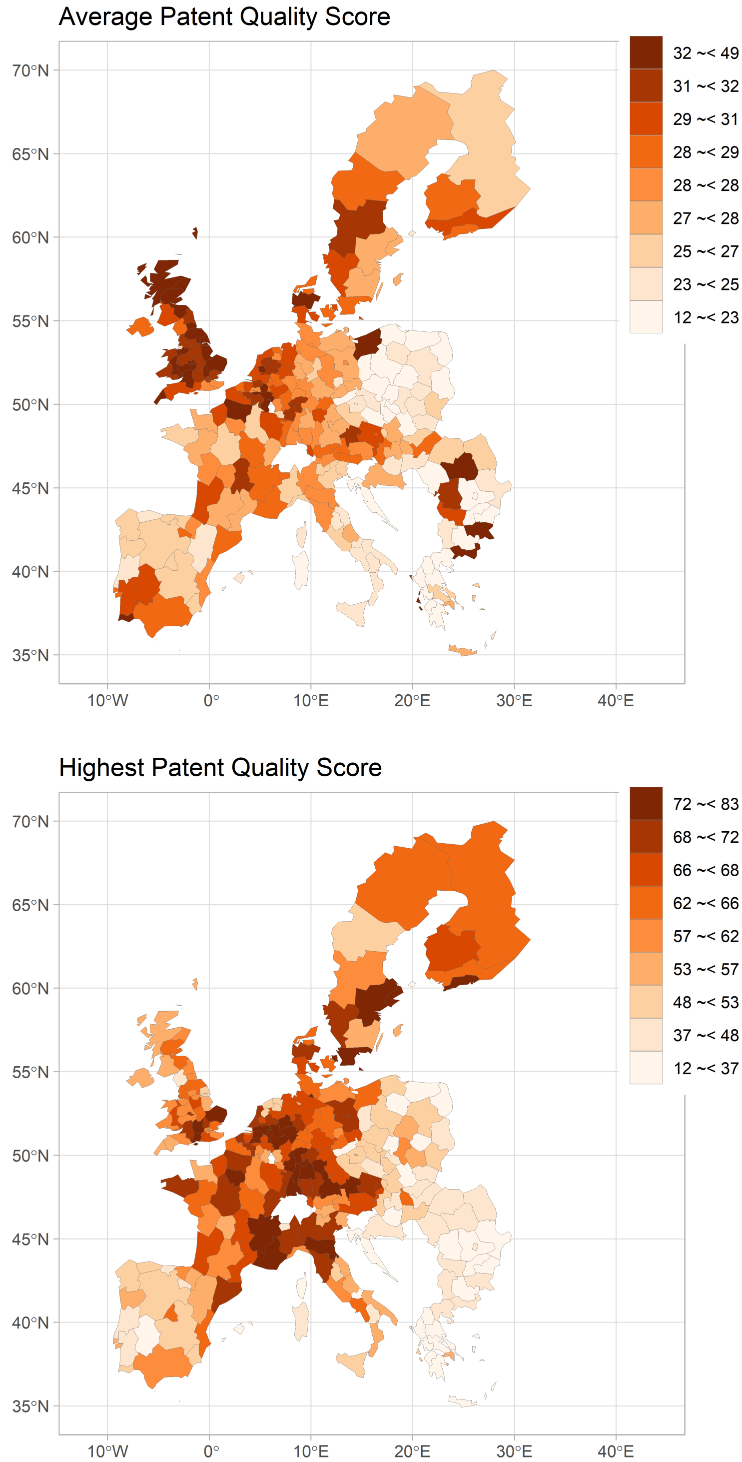


Figure 3: Maps of Technological Complexity

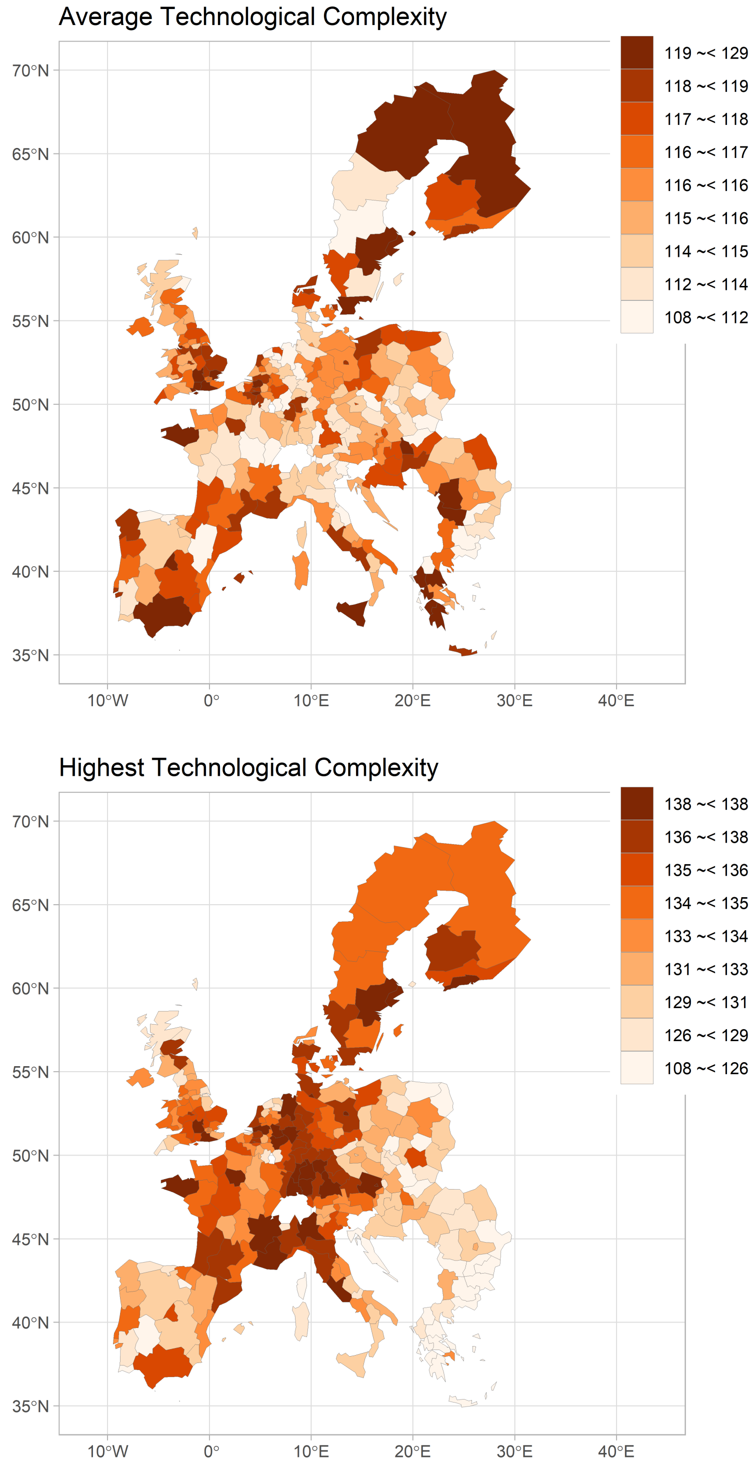
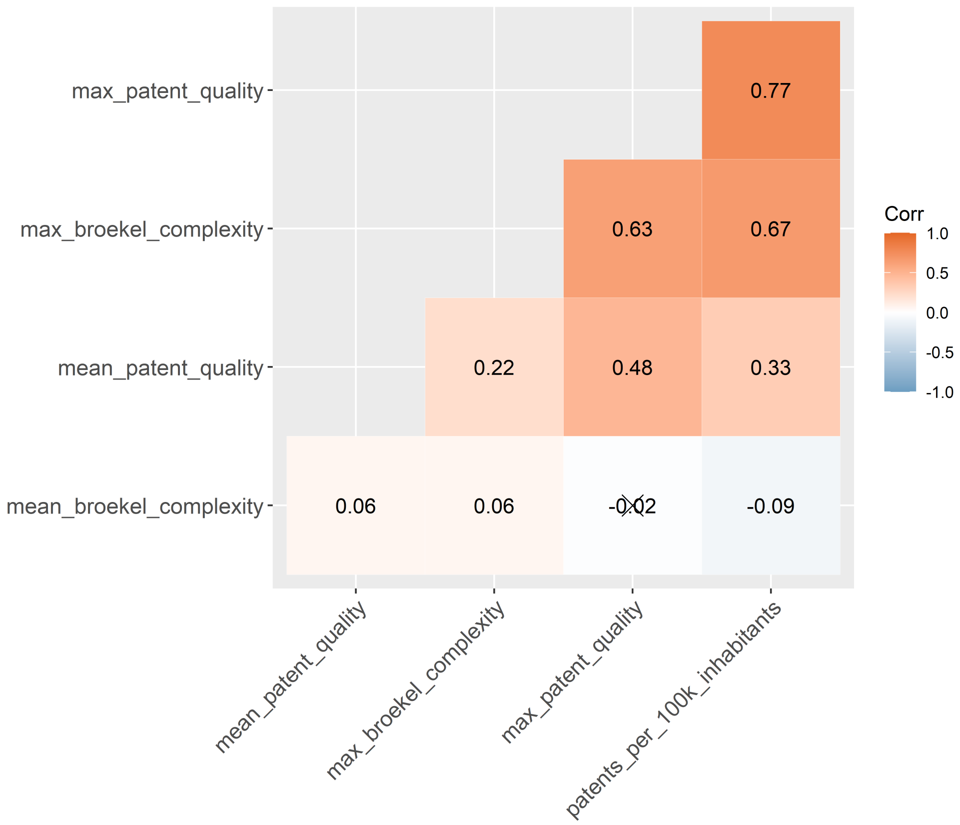


Figure 4: Correlation Matrix of Patent Variables



5.2 Corruption Risk

Looking at the map of Single Bidding and the EQI Corruption Pillar, an east-west gap is clearly visible. Greece could be seen as an outlier when it comes to Single Bidding. According to the data, it is the country with the least occurrence of Single Bidding in competitive markets. On the other hand, the EQI shows that according to its citizens' perception, it is among the most corrupt countries in Europe. This does not necessarily have to mean that the data is erroneous. Single Bidding is, after all, simply an indirect proxy for corruption risk in public procurement. It can, of course, be that either the Greece public procurement is indeed not very corrupt or that other corrupt methods are used that are not reflected in the Single Bidding red flag. Nonetheless, at least one of the two corruption proxies does not do a good job of reflecting the true extent of corruption in Greece. Therefore, the Greek regions will be excluded from the analysis to avoid distorted results.

Figure 5: Map of Single Bidding

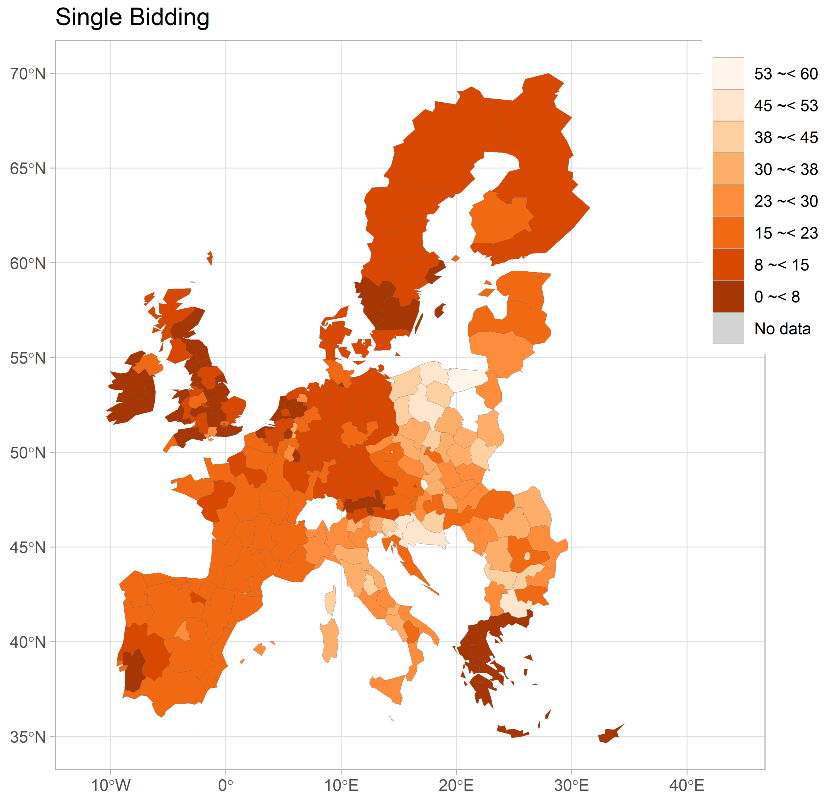
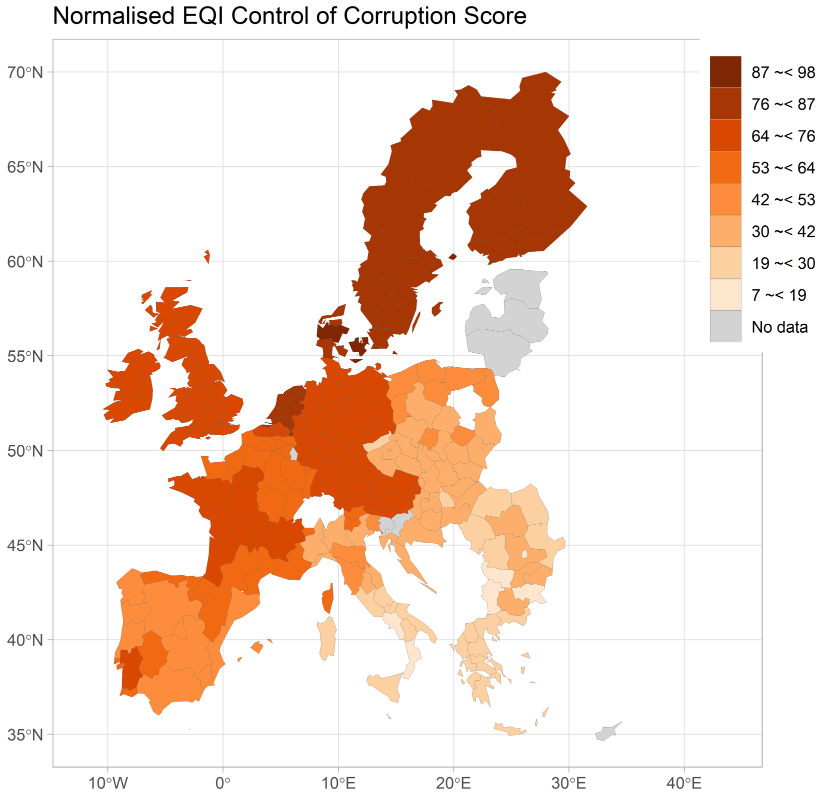


Figure 6: Map of EQI Corruption Pillar



Besides the spatial distribution of corruption in Europe, it is, of course, also interesting to see how it has developed over the period analysed. The answer is quite alarming. While the ratio of Single Bidding contracts has risen significantly since 2014, the average EQI has dropped by several points. Both corruption proxies show that corruption in the EU has increased quite substantially from 2011 to 2017.

5.3 Sources of Variation

Because the data are panel data, one can measure cross-sectional variation between regions and variation over time within regions. I calculated the average coefficient of variation[1] for the corruption and innovation proxies. Table 5 shows what could already be expected. The variation within regions is considerably smaller than between regions for all of the variables. Nev-

[1] Ratio of the standard deviation to the mean times 100

Figure 7: Development of Single Bidding over Time

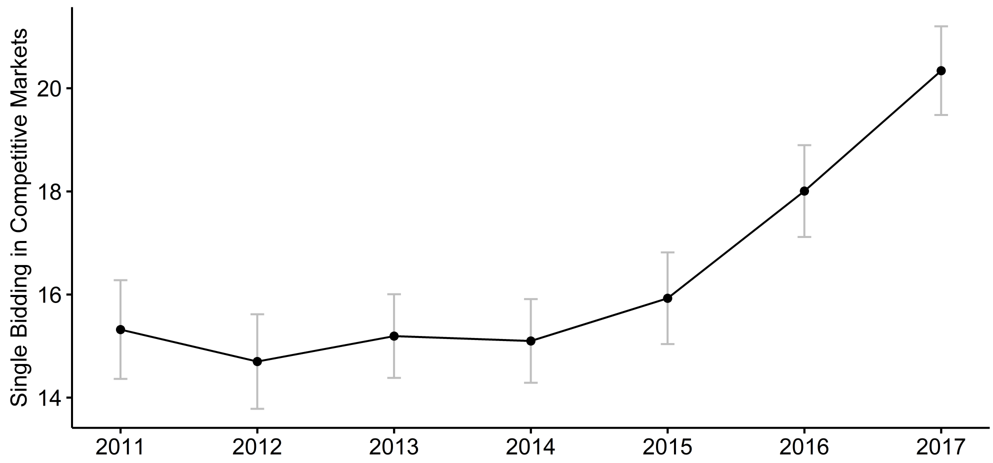
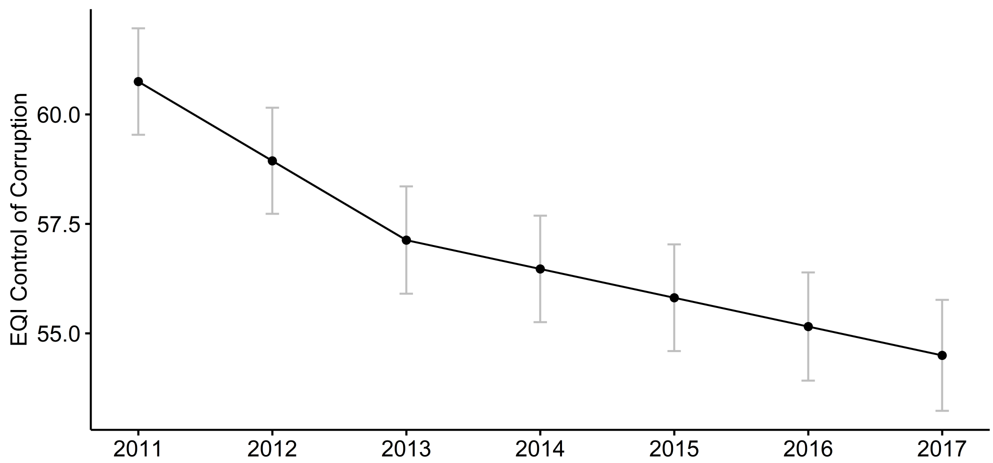


Figure 8: Development of the EQI over Time



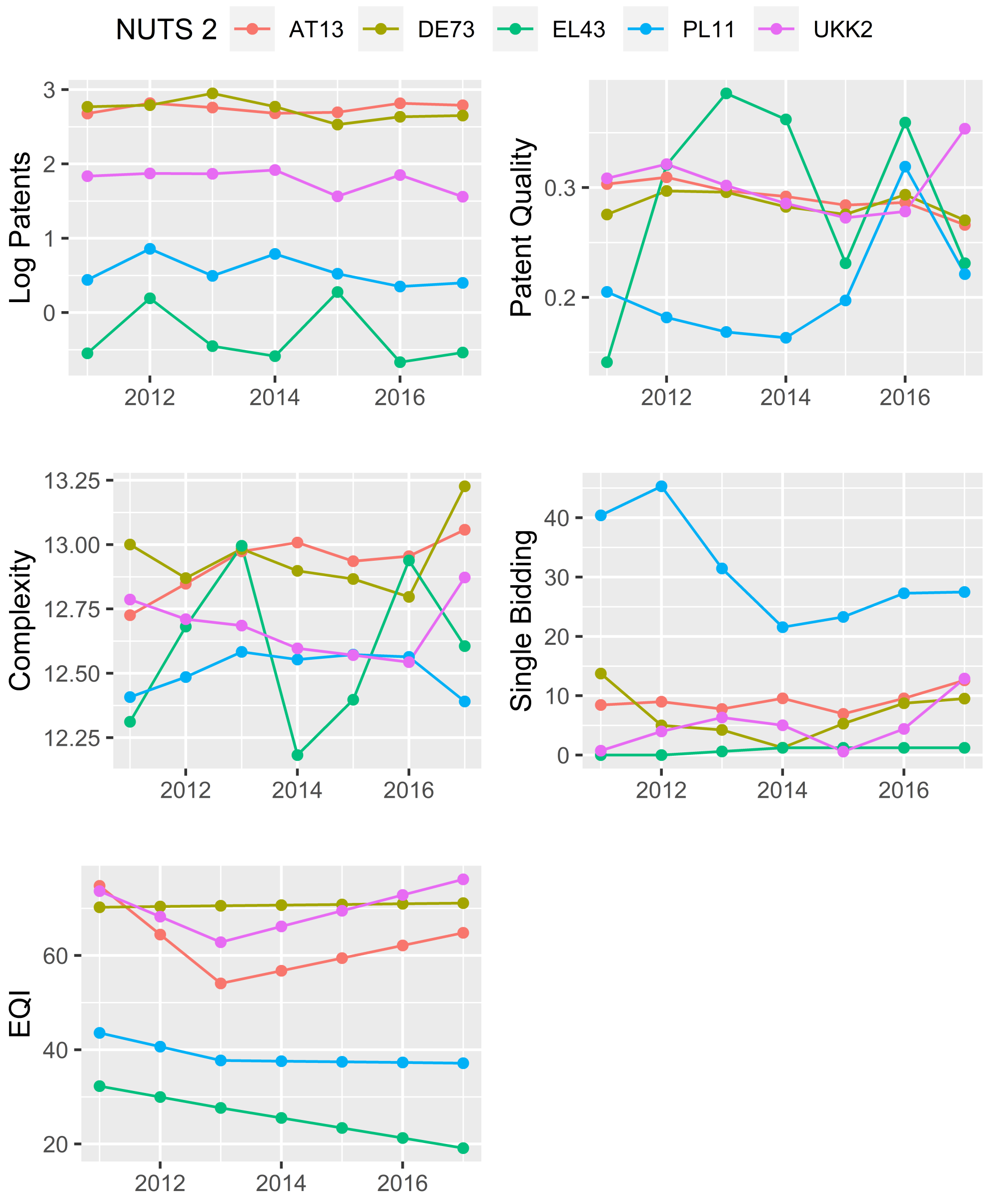
ertheless, there is variation in both dimensions. This is confirmed by the pvar function, implemented in the R-package *plm*, that tests for statistically significant variation.

Figure 9 helps in getting a better intuition regarding the sources of variation. Five regions are randomly selected, and the respective variation over time of the innovation and corruption proxies plotted. In this sample, the log-transformed number of patent applications per 100,000 inhabitants shows large between region heterogeneity but is rather stable within regions. The regions with high patent counts are more stable over time than those with little patenting activity, at least in relative terms. The average patent quality shows a similar picture. While the Austrian, German, and British regions do not vary much among themselves and over time, the Polish and Greek regions vary considerably over time. This is not surprising, as individual patents make a larger difference in the mean when there is only a small number of patents overall. The average technological complexity is relatively homogeneous amongst the sample regions except for the German region Kassel which clearly shows an upwards trend and reaches a significantly higher average technological complexity by 2017. The Single Bidding plot shows again the extraordinarily low Single Bidding ratio in Greek regions. The Polish region Łódzkie on the other hand has a significantly higher percentage of Single Bidding contracts and strong variation over time. The final plot of the EQI reveals no surprising patterns. There are large between region differences, but regions change rather slowly over time.

Table 5: Variation Within and Between Regions

	Within CV	Between CV
Log Patents	20.21	45.96
Patent Quality	12.58	19.99
Complexity	2.06	3.38
Single Bidding	56.85	90.24
EQI	9.59	34.08

Figure 9: Plots of Variation



6 Methods

Choosing an appropriate research method and model specification is not trivial, especially in panel data analyses. The two-dimensionality[2] of panels can potentially enhance the quality of estimates [59], However, it gets quite complex when it comes to the exact specifications due to the variety of applicable estimators and model specifications and the ongoing discussion in academia regarding the state-of-the-art [8]. Often, there is no one best approach, and in the end, it comes down to the researcher's preferences and beliefs [46]. This section outlines the advantages and disadvantages of commonly used model types and concludes with my selection.

The most basic approach to panel regressions is to apply ordinary least squares (OLS) to the pooled data. Problems can arise due to the assumption that the error term is not correlated with the independent variables because this would require that all influencing factors that are specific to the individual units are included as control variables in the model [58]. Pooled OLS models are therefore insufficient when the observations are not independent of each other, which is given in a real panel structure [59]. Furthermore, a straightforward interpretation of the estimated coefficients is impossible because within-unit and between-units effects are mixed [7]. Another rather simple model type is the between estimator. The data is simply averaged over the years, completely eliminating the time dimension. Regular OLS is then applied to the transformed, cross-sectional data. The main advantage of this model type is that noise in the data is significantly reduced, albeit at a loss of information.

A widely used model type that partially accounts for omitted variable bias is the fixed effects model. The major advantage is that unobserved heterogeneity is controlled for in the form of effects that are specific to each unit and constant over time. Additionally, shocks that affect all units equally can be modelled via time-fixed effects [58]. An easy to understand implementation of this model type is the Least Squares Dummy Variable model (LSDV). As its name implies, i-1 dummy variables are included in the formula to estimate the individual-fixed effects. The same concept can be applied to time-variant shocks that are identical for all units. In this case, t-1 time dummies are included in the model. This model type assumes constant slopes but intercepts that vary between the units. By including unit dummy variables, only variation within units, in our case regions, is considered. Because of this, it is also called within estimator. One major drawback of the fixed effects model is that a large number of cross-sectional units can lead to an insufficient number

[2]individuals and time

of degrees of freedom which impedes powerful statistical tests. Further issues regarding inflated standard errors can arise due to multicollinearity [59]. Additionally, time-invariant variables are fully absorbed by the fixed effects. Their effect can therefore not be estimated because only the variation within a region is analysed [7]. This could potentially be problematic considering the rather short time period for which regional corruption data is available and the fact that the levels of corruption and innovation are rather persistent and do not change rapidly, especially not on a yearly basis [28].

If applicable, the random effect (RE) models are more efficient and allow for the inclusion of time-invariant variables. In the RE model, the intercept is treated as a random outcome variable. More precisely, it is assumed to be a function of a mean value plus a random error. This random error is time-invariant and specific to a particular observation [59].

Autocorrelation, or serial correlation, is another potential factor that needs to be considered in the model selection process. One way to deal with serial correlation is to include the lagged dependent variable as a regressor. This type of models is called dynamic panel models [59]. Kelly and Keele (2004) use a Monte Carlo simulation to show that excluding the lagged dependent variable when autocorrelation is present leads to a strong positive bias in the estimated coefficients. On the other hand, if the lagged dependent variable is wrongly included in a static model, the coefficients of the predictors can be heavily underestimated [37]. Achen (2000) shows that lagged dependent variables that are included as regressors often take on highly significant, large coefficients while suppressing the effects of the other explanatory variables. Occasionally, even the signs of the other coefficients are flipped. To demonstrate this, he takes a regression with social welfare expenditures as the dependent variable and the percentage of inhabitants older than 64 and the unemployment rate as explanatory variables. He shows that before the inclusion of the lagged dependent variable, the explanatory variables take on sensible and significant coefficient values. However, after the inclusion, the results shift tremendously, with a collapse of the estimated coefficients to about 1 % and 2 % of their original values. The coefficient of unemployment even becomes negative, which would mean that a higher unemployment rate leads to lower social welfare expenditure [3]. Furthermore, including a lagged term of the dependent variable in fixed effects models causes the so-called Nickell bias [47].

According to the Hausman test, the RE model is no suitable option for these data and is therefore not used. Due to the small number of available years in the data and the finite-sample bias that occurs when lagged dependent variables are included in fixed effects models, I choose a static specification. Furthermore, I model the variation between regions and within regions over time separately with the between and within estimators, respectively. This separation allows for a clearer interpretation of the results because the between and within region effects are not averaged in one model [38].

7 Results

7.1 Correlations

Before presenting the regression results, I want to start this section with the simplest of statistical relationships, namely correlations. The correlations between the corruption and innovation proxies show expected and unexpected relations. As anticipated, the Single Bidding red flag is negatively correlated with all innovation proxies, the strongest correlation being with the patent applications, weighted by their respective quality. The correlation between the corruption pillar of the EQI with the innovation proxies is even stronger, with an impressive correlation coefficient of 0.70 with the number of patent applications per 100,000 inhabitants. What is rather surprising is that Single Bidding seems to be the only red flag that is negatively correlated with the innovation proxies. The other three red-flag indicators are mildly positively correlated with the patenting activity, which results in the CRI[3], being basically uncorrelated with the patent variables.

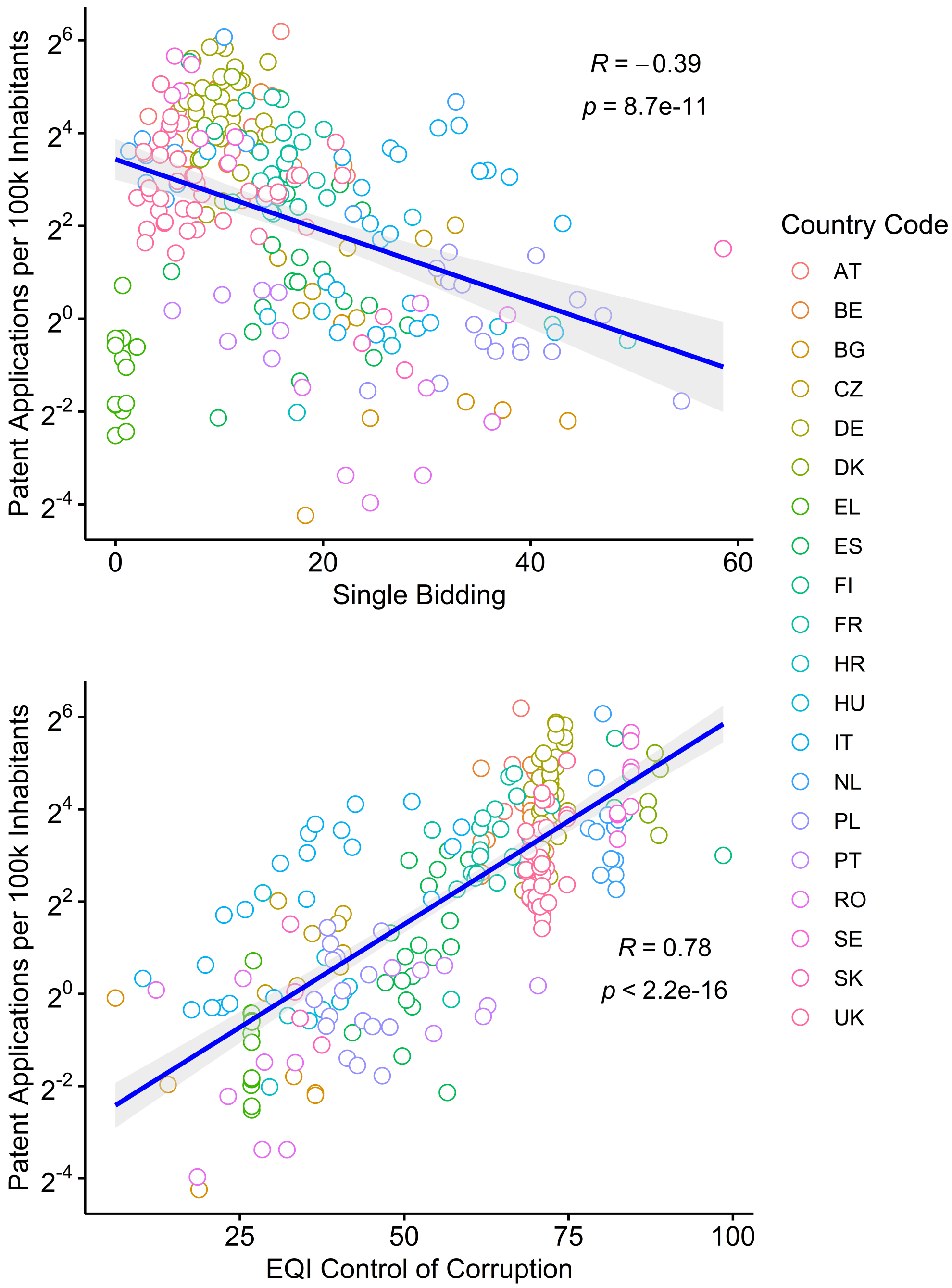
These results, in combination with the fact that using just Single Bidding instead of the CRI as a proxy for corruption risk in public procurement is suggested and validated by the authors who have also proposed the CRI, motivates the choice to continue with Single Bidding together with the EQI corruption pillar as corruption proxies for the rest of this thesis [25].

Table 6: Correlations Patenting and Corruption

	Quantity	Quality	Complexity	Weighted
CRI	-0.02	0.03	-0.03	-0.04
No Call	0.13	0.16	0.03	0.10
Non Open	0.03	0.05	0.02	0.01
Single Bidding	-0.36	-0.30	-0.21	-0.38
Tax Haven	0.06	0.05	0.08	0.07
EQI Corruption	0.70	0.53	0.39	0.69

[3] the arithmetic mean of the red flags indicators

Figure 10: Scatterplots - Patent Quantity and Corruption Proxies



7.2 Effect on Patent Quantity

7.2.1 Between Estimator

The between region analysis of the effect of corruption on the quantity of patent applications results in two main regression specifications. The first has the following form:

$$\overline{Patents_r} = \alpha + \beta_1 \overline{Corrupt_r} + \gamma \overline{X}_r + \overline{\epsilon}_r \tag{1}$$

where α is the intercept and $\overline{Patents_r}$ represents the number of patent applications per 100,000 inhabitants. $\overline{Corrupt_r}$ is a matrix containing Single Bidding and the EQI Corruption Pillar. $\overline{X}_i$ is the matrix of control variables, so GDP per capita in purchasing power standards (PPS), the unemployment rate, population density, the share of workers in the primary sector and the rate of inhabitants who completed the tertiary education level, with γ being the vector containing the respective coefficients. $\overline{\epsilon}_i$ is the error term. All variables but the EQI are transformed by applying the natural logarithm. The bar above the variables shows that they represent averages over all years for which data are available. A constant of 1 is added to the Single Bidding rate to avoid the problem of negative infinite values when there were no Single Bidding contracts at all in a given year and region.

With the second specification, I control directly for the most important innovation inputs. The purpose of this specification is to see whether corruption has a statistically significant effect, even when the level of innovation inputs is held constant. This corresponds to the concept of innovation efficiency [12]. To be precise, the model has the following form:

$$\overline{Patents_r} = \alpha + \beta_1 \overline{Corrupt_r} + \gamma \overline{X}_r + \omega \overline{Inputs_r} + \overline{\epsilon}_r \tag{2}$$

where the newly introduced matrix $\overline{Inputs_i}$ contains the percentage of employees working in high-technology sectors and the intramural expenditure on research and development.

Table 7 shows the regression results. In the specification without innovation inputs, Single Bidding as a proxy for corruption in the public procurement has a negative estimated effect on the number of patent applications per 100,000 inhabitants at the 0.1 significance level. An increase in the rate of Single Bidding contracts by 10% is associated with a decrease of patent applications per capita of 1.46%. The control of corruption pillar of the EQI has a highly significant positive effect that only becomes slightly smaller when the innovation inputs are included, with an increase of one standard deviation being associated with an increase in patent applications per capita of 0.446%.

7.2.2 Within Estimator

The fixed effects regressions resemble the previously described regression formulae in most parts. However, due to the exploitation of the time-series dimension of the data, some key distinctions are apparent. The regressions have the following forms:

$$PatentsMA_{i,t} = \alpha_r + \psi_t + \beta_1 CorruptMA_{r,t} + \gamma X_{r,t} + \epsilon_{r,t} \tag{3}$$

$$PatentsMA_{r,t} = \alpha_r + \psi_t + \beta_1 CorruptMA_{r,t} + \gamma X_{r,t} + \omega Inputs_{r,t} + \epsilon_{r,t} \tag{4}$$

where α_r captures region-fixed effects, ψ_t year-fixed effects and $PatentsMA_{r,t}$ is a left-aligned three year moving average of the form:

$$\frac{1}{3}\sum_{i=t}^{t+2} Patents_{r,i} \tag{5}$$

This filter is applied to bring the innovation proxy closer to the true level of innovation in a given year t. Patent applications reflect innovative activity that predates the filing of the patent application by several years [4]. It is therefore sensible to consider not only patent applications in t but also in $t+1$ and $t+2$ in the approximation of the innovation level in year t. $CorruptMA_{r,t}$ is a right-aligned three year moving average of the form:

$$\frac{1}{3}\sum_{i=t-2}^{t} Corrupt_{r,i} \tag{6}$$

The primary reason for smoothing, in this case, is to reduce the negative effect of noise in the Single Bidding variable. Contrary to the approach used for smoothing the number of patent applications, the moving average is calculated by considering the previous two years' values. The reason for this is to better model the environment in which the invention of new technologies has taken place and the fact that corruption is expected to affect the level of innovation rather delayed than immediately. To demonstrate this intuitively, imagine being the director of a research facility. It will most likely not be possible to relocate the facility immediately in response to a changed level of corruption because this requires quite some lead time. However, in the longer term, resources will presumably be allocated to more favourable locations.

In the within estimation, Single Bidding has a highly significant negative estimated effect, while the effect of the EQI is not significantly different from

0. This is an interesting observation, considering that the two corruption proxies are based on very different concepts with very different advantages and disadvantages. One of the main advantages of the Single Bidding indicator is that it is not based on perceptions and therefore does not suffer from the stickiness associated with perception-based indicators, like the EQI, for example. The fact that the EQI is highly significant in the between but insignificant in the within estimation and vice versa for Single Bidding does not allow for a causal conclusion to be drawn about the specific reason. Still, it is an interesting finding that could be the subject of future research. The estimated effect size of Single Bidding is arguably rather small. An increase of the Single Bidding rate of 10 % is associated with an expected decrease in the number of patents per capita of 0.29%, ceteris paribus.

Table 7: Patent Quantity - Between Estimator

	Dependent variable:			
	LN Patent Applications per 100,000 Inhabitants			
	(1)	(2)	(3)	(4)
Single Bidding	-0.146^{*}	-0.052		
	(0.084)	(0.068)		
EQI Corruption			0.577^{***}	0.446^{***}
			(0.073)	(0.063)
GDP PPS	2.071^{***}	0.331	1.745^{***}	0.255
	(0.193)	(0.202)	(0.178)	(0.184)
Unemployment	-0.646^{***}	-0.597^{***}	-0.453^{***}	-0.410^{***}
	(0.109)	(0.086)	(0.099)	(0.080)
Pop. Density	-0.284^{***}	-0.095^{**}	-0.160^{***}	-0.032
	(0.053)	(0.044)	(0.050)	(0.041)
Primary Sector	-0.484^{***}	-0.202^{***}	-0.401^{***}	-0.130^{*}
	(0.087)	(0.073)	(0.078)	(0.067)
Tertiary Ed.	0.154	0.007	-0.422^{**}	-0.525^{***}
	(0.197)	(0.160)	(0.190)	(0.160)
Hightech		-0.210^{*}		0.080
		(0.124)		(0.117)
R&D Spending		0.821^{***}		0.704^{***}
		(0.064)		(0.060)
Constant	-16.240^{***}	-4.140^{**}	-12.650^{***}	-2.285
	(2.089)	(1.882)	(1.887)	(1.700)
Observations	246	246	246	246
R^2	0.730	0.841	0.784	0.869
Adjusted R^2	0.723	0.836	0.778	0.864
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 8: Patent Quantity - Within Estimator

	Dependent variable:			
	LN Patent Applications per 100.000 Inhabitants			
	(1)	(2)	(3)	(4)
Single Bidding	-0.028^{***}	-0.029^{***}		
	(0.010)	(0.010)		
EQI Corruption			-0.037	-0.039
			(0.030)	(0.030)
GDP PPS	0.238	0.221	0.277^{*}	0.259^{*}
	(0.153)	(0.154)	(0.153)	(0.154)
Unemployment	-0.022	-0.020	-0.022	-0.020
	(0.030)	(0.030)	(0.030)	(0.030)
Population Density	-1.531^{***}	-1.438^{***}	-1.578^{***}	-1.488^{***}
	(0.300)	(0.300)	(0.301)	(0.301)
Primary Sector	-0.087^{***}	-0.083^{***}	-0.086^{***}	-0.082^{***}
	(0.030)	(0.030)	(0.030)	(0.030)
Tertiary Ed.	0.313^{***}	0.265^{***}	0.315^{***}	0.268^{***}
	(0.068)	(0.069)	(0.069)	(0.070)
Hightech		0.128^{***}		0.124^{***}
		(0.037)		(0.037)
R&D Spending		0.017		0.021
		(0.027)		(0.027)
Observations	1,722	1,722	1,722	1,722
Note:				*p<0.1; **p<0.05; ***p<0.01

7.3 Effect on Patent Quality

The previous section examined, whether corruption affects innovation performance, as measured by the quantity of patent applications. However, the number of patent applications alone is a rather insufficient proxy for the innovation performance of a region. To get a broader picture of this matter, the quality dimension is also considered. In this section, I analyse the effect of corruption on patent quality through regressions of the following form:

$$\overline{Quality}_r = \alpha + \beta_1 \overline{Corrupt}_r + \gamma \overline{X}_r + \overline{\epsilon}_r \tag{7}$$

$$\overline{Quality}_r = \alpha + \beta_1 \overline{Corrupt}_r + \gamma \overline{X}_r + \omega \overline{Inputs}_r + \overline{\epsilon}_r \tag{8}$$

$$\overline{Quality}_r = \alpha + \beta_1 \overline{Corrupt}_r + \gamma \overline{X}_r + \omega \overline{Inputs}_r + \beta_2 \overline{Patents}_r + \overline{\epsilon}_r \tag{9}$$

$$QualityMA_{i,t} = \alpha_r + \psi_t + \beta_1 CorruptMA_{r,t} + \gamma X_{r,t} + \epsilon_{r,t} \tag{10}$$

$$QualityMA_{r,t} = \alpha_r + \psi_t + \beta_1 CorruptMA_{r,t} + \gamma X_{r,t} + \omega Inputs_{r,t} + \epsilon_{r,t} \tag{11}$$

$$QualityMA_{r,t} = \alpha_r + \psi_t + \beta_1 CorruptMA_{r,t} + \gamma X_{r,t} + \omega Inputs_{r,t} + \beta_2 Patents_{r,t} + \epsilon_{r,t} \tag{12}$$

To avoid redundancy, I will only describe the changes compared to the previously reported regression formulae. $QualityMA_{r,t}$ denotes the average patent quality of the top 10% highest-quality patents in a region and year. Patent quality in general is measured by the OECD patent quality index as described in more detail in the data section of this thesis. The variable is z-score normalized to facilitate the coefficient interpretation. A third specification for the between and within estimation is introduced, in which I control for the number of patent applications per capita, $Patents_{r,t}$. This is done to test whether a potential increase in patent quality is only indirectly induced by an increased number of patent applications or whether corruption has a significant effect on the quality of patents, even when the quantity is held constant.

In the between estimation, Single Bidding has a highly significant negative effect on patent quality, even in the strictest specification in which the quantity of patents is controlled for. For the EQI, the highly significant positive effect vanishes when the number of patents per capita is included as an explanatory variable. The fixed effects regression shows quite a similar picture. Single Bidding retains its significant negative effect, albeit to a smaller degree. In the strictest between regression, a 1% increase of the Single Bidding rate is associated with a decrease of patent quality of -0.150 standard deviations, and in the within estimation -0.063 standard deviations, respectively. In the fixed effects model, the EQI Corruption Pillar does not lose its significant positive effect when the quantity of patents is controlled for. On the contrary, it even retains its effect size of 0.486. Interestingly, the corruption proxies retain their sign, regardless of whether only the variation between or within regions is considered. This is not the case for some of the control variables. If we look at the population density, for example, the estimated effect in the between estimation is positive, signalling that more urban regions with higher population density produce higher-quality patents. However, within a given region, an increasing population density is associated with a rather strong negative effect on the quality of patents, ceteris paribus.

Table 9: Patent Quality - Between Estimator

	Dependent variable:					
	Avg. Patent Quality in 90^{th} Percentile					
	(1)	(2)	(3)	(4)	(5)	(6)
Single Bidding	−0.207***	−0.168***	−0.150***			
	(0.061)	(0.059)	(0.054)			
EQI Corruption				0.280***	0.205***	0.064
				(0.058)	(0.059)	(0.061)
GDP PPS	0.534***	−0.305*	−0.416**	0.417***	−0.319*	−0.399**
	(0.141)	(0.174)	(0.161)	(0.141)	(0.173)	(0.163)
Unemployment	−0.129	−0.100	0.099	−0.083	−0.070	0.059
	(0.079)	(0.074)	(0.075)	(0.078)	(0.075)	(0.075)
Pop. Density	0.026	0.116***	0.147***	0.085**	0.149***	0.159***
	(0.039)	(0.038)	(0.035)	(0.040)	(0.038)	(0.036)
Primary Sector	−0.136**	0.005	0.072	−0.100	0.027	0.068
	(0.063)	(0.063)	(0.059)	(0.062)	(0.063)	(0.060)
Tertiary Ed.	0.221	0.137	0.134	0.024	−0.004	0.162
	(0.144)	(0.137)	(0.127)	(0.151)	(0.150)	(0.144)
Hightech		−0.064	0.006		−0.008	−0.033
		(0.107)	(0.099)		(0.110)	(0.103)
R&D Spending		0.391***	0.117*		0.355***	0.133**
		(0.055)	(0.066)		(0.057)	(0.067)
Patent Quantity			0.334***			0.316***
			(0.052)			(0.057)
Constant	−5.351***	0.520	1.904	−4.547***	0.526	1.249
	(1.520)	(1.618)	(1.509)	(1.498)	(1.594)	(1.510)
Observations	246	246	246	246	246	246
R^2	0.441	0.541	0.611	0.467	0.548	0.600
Adjusted R^2	0.427	0.526	0.596	0.454	0.533	0.584
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table 10: Patent Quality - Within Estimator

	Dependent variable:					
	Avg. Patent Quality in 90th Percentile					
	(1)	(2)	(3)	(4)	(5)	(6)
Single Bidding	−0.067**	−0.065**	−0.063*			
	(0.033)	(0.032)	(0.033)			
EQI Corruption				0.495***	0.485***	0.486***
				(0.095)	(0.096)	(0.096)
GDP PPS	0.019	−0.067	−0.097	−0.073	−0.110	−0.144
	(0.491)	(0.495)	(0.495)	(0.487)	(0.491)	(0.491)
Unemployment	−0.029	−0.026	−0.022	0.052	0.051	0.054
	(0.095)	(0.095)	(0.095)	(0.095)	(0.095)	(0.095)
Pop. Density	−0.243	−0.379	−0.197	−1.083	−1.223	−1.022
	(0.963)	(0.966)	(0.975)	(0.957)	(0.960)	(0.969)
Primary Sector	−0.001	−0.008	−0.001	0.025	0.017	0.025
	(0.096)	(0.096)	(0.096)	(0.095)	(0.095)	(0.096)
Tertiary Ed.	−0.582***	−0.507**	−0.542**	−0.322	−0.243	−0.282
	(0.220)	(0.223)	(0.225)	(0.221)	(0.225)	(0.226)
Hightech		−0.265**	−0.271**		−0.268**	−0.274**
		(0.118)	(0.118)		(0.117)	(0.117)
R&D Spending		0.112	0.110		0.053	0.051
		(0.086)	(0.086)		(0.086)	(0.086)
Patent Quantity			0.073			0.079
			(0.054)			(0.054)
Observations	1,722	1,722	1,722	1,722	1,722	1,722

Note: *p<0.1; **p<0.05; ***p<0.01

7.4 Effect on Technological Complexity

So far, the effects of corruption on the quantity and quality of patents were examined. However, the question remains whether this effect is uniform across all technologies or whether the technological complexity of innovations is affected as well. The regression specifications are identical to the ones used in the previous subsection, with the only difference being that the dependent variable is now the average technological complexity of the top 10% most technologically complex patents, as measured through the structural diversity of the respective technology. To facilitate the interpretation of results, the variable was z-score normalized.

In the between estimation, I do not find any statistically significant results of corruption on the technological complexity of a region, regardless of which of the two proxies and which control variables are considered.

In the fixed effects model, however, Single Bidding has a highly significant negative effect across all specifications. The EQI Corruption Pillar has a positive estimated effect that is statistically significant at the 0.05 level without the direct innovation inputs and the 0.1 level when direct innovation inputs and the number of patent applications per capita are included in the regression. In the model with all controls included, the estimated effect of Single Bidding is -0.091, so an increase in the Single Bidding rate by 10% is expected to lead to a decrease of the technological complexity of a region by -0.91 standard deviations, ceteris paribus.

What strikes the eye is the exceptionally strong negative effect of population density. Additional research that would, unfortunately, exceed the scope of this thesis would be needed to explain this relationship in more detail.

Table 11: Technological Complexity - Between Estimator

	Dependent variable:					
	Avg. Technological Complexity in 90^{th} Percentile					
	(1)	(2)	(3)	(4)	(5)	(6)
Single Bidding	0.035	0.017	0.018			
	(0.084)	(0.082)	(0.082)			
EQI Corruption				0.038	0.034	0.005
				(0.080)	(0.081)	(0.090)
GDP PPS	0.547^{***}	-0.474^{**}	-0.497^{**}	0.516^{***}	-0.480^{**}	-0.496^{**}
	(0.190)	(0.237)	(0.239)	(0.196)	(0.238)	(0.239)
Unemployment	-0.119	-0.012	0.029	-0.089	0.011	0.038
	(0.109)	(0.103)	(0.113)	(0.109)	(0.103)	(0.109)
Pop. Density	0.002	0.086^{*}	0.093^{*}	0.011	0.091^{*}	0.093^{*}
	(0.053)	(0.051)	(0.052)	(0.055)	(0.053)	(0.053)
Primary Sector	-0.115	0.127	0.141	-0.106	0.135	0.143
	(0.086)	(0.086)	(0.088)	(0.087)	(0.087)	(0.088)
Tertiary Ed.	0.190	-0.086	-0.088	0.121	-0.144	-0.111
	(0.199)	(0.192)	(0.192)	(0.210)	(0.206)	(0.211)
Hightech		0.415^{***}	0.431^{***}		0.449^{***}	0.444^{***}
		(0.147)	(0.148)		(0.151)	(0.151)
R&D Spending		0.419^{***}	0.364^{***}		0.406^{***}	0.361^{***}
		(0.076)	(0.099)		(0.078)	(0.098)
Patent Quantity			0.067			0.064
			(0.076)			(0.084)
Constant	-5.923^{***}	1.637	1.930	-5.416^{***}	1.875	2.022
	(2.044)	(2.188)	(2.215)	(2.085)	(2.196)	(2.206)
Observations	246	246	246	246	246	246
R^2	0.177	0.317	0.319	0.177	0.317	0.319
Adjusted R^2	0.156	0.293	0.293	0.156	0.294	0.293
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table 12: Technological Complexity - Within Estimator

	Dependent variable:					
	Avg. Technological Complexity in 90^{th} Percentile					
	(1)	(2)	(3)	(4)	(5)	(6)
Single Bidding	-0.094^{***}	-0.098^{***}	-0.091^{***}			
	(0.035)	(0.035)	(0.035)			
EQI Corruption				0.199^{*}	0.176^{*}	0.180^{*}
				(0.104)	(0.105)	(0.104)
GDP PPS	0.999^{*}	0.834	0.705	1.028^{*}	0.892^{*}	0.753
	(0.533)	(0.537)	(0.533)	(0.534)	(0.537)	(0.533)
Unemployment	-0.067	-0.053	-0.037	-0.021	-0.010	0.005
	(0.103)	(0.103)	(0.102)	(0.104)	(0.104)	(0.103)
Pop. Density	-4.956^{***}	-4.672^{***}	-3.899^{***}	-5.522^{***}	-5.237^{***}	-4.429^{***}
	(1.047)	(1.048)	(1.048)	(1.049)	(1.051)	(1.051)
Primary Sector	0.185^{*}	0.197^{*}	0.228^{**}	0.203^{*}	0.213^{**}	0.245^{**}
	(0.105)	(0.104)	(0.104)	(0.105)	(0.105)	(0.104)
Tertiary Ed.	0.256	0.118	-0.033	0.406^{*}	0.267	0.110
	(0.239)	(0.242)	(0.242)	(0.242)	(0.246)	(0.245)
Hightech		0.312^{**}	0.286^{**}		0.301^{**}	0.275^{**}
		(0.128)	(0.127)		(0.128)	(0.127)
R&D Spending		0.191^{**}	0.182^{**}		0.168^{*}	0.159^{*}
		(0.093)	(0.092)		(0.094)	(0.093)
Patent Quantity			0.312^{***}			0.318^{***}
			(0.058)			(0.058)
Observations	1,722	1,722	1,722	1,722	1,722	1,722
Note:				*p<0.1; **p<0.05; ***p<0.01		

7.5 Effect on Technological Composition

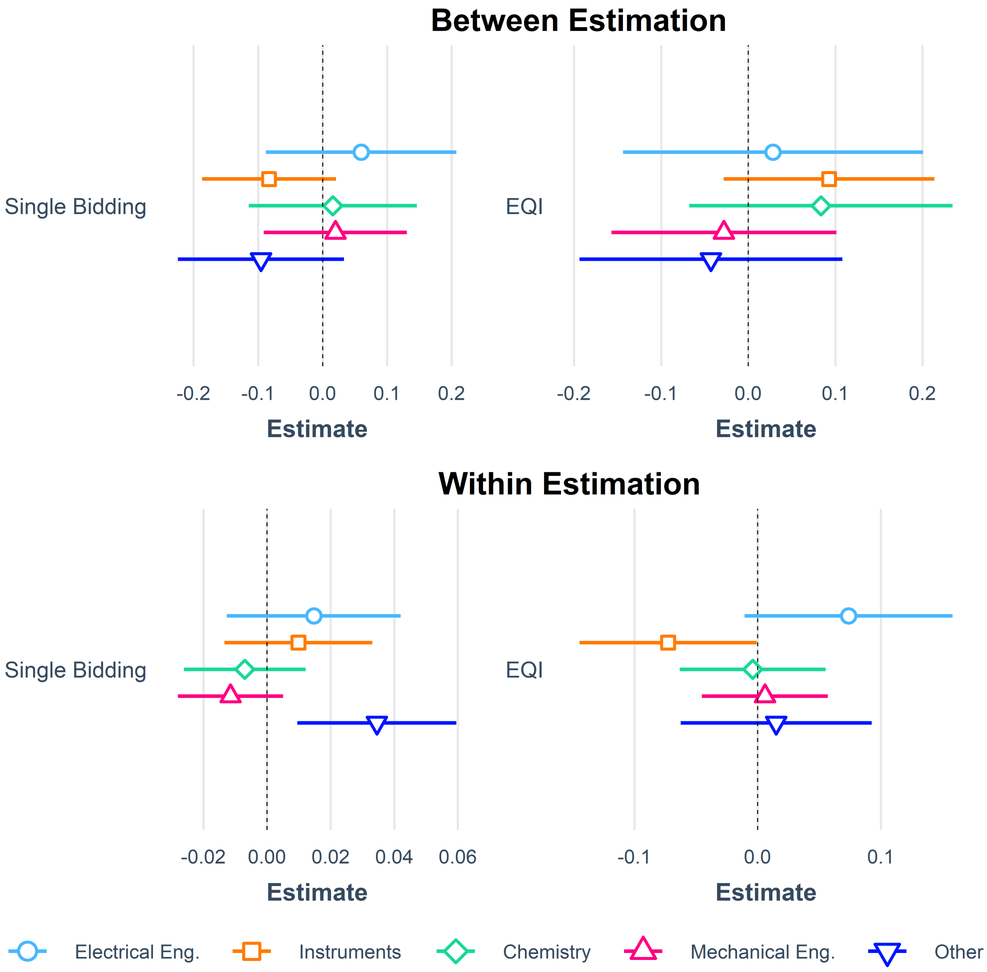
The final research question that is dealt with in this thesis is whether the effect of corruption is homogeneous across industries or not. To analyse this, patents are classified according to the WIPO IPC - Technology Concordance Table into five sectors, namely *Electrical Engineering*, *Instruments*, *Chemistry*, *Mechanical Engineering* and *Other Fields*. Regressions of the following form were constructed to estimate the effect of corruption on the share of patents belonging to each of these sectors:

$$\overline{Sector}_r = \alpha + \beta_1 \overline{Corrupt}_r + \gamma \overline{X}_r + \beta_2 \overline{Inputs}_r + \beta_3 \overline{Patents}_r + \overline{\epsilon}_r \tag{13}$$

$$Sector_{r,t} = \alpha_r + \psi_t + \beta_1 CorruptMA_{r,t} + \gamma X_{r,t} + \beta_2 Inputs_{r,t} + \beta_3 Patents_{r,t} + \epsilon_{r,t} \tag{14}$$

where *Sector* is a matrix containing the share of patent applications belonging to each of the five sectors in relation to the total patent applications in a region. Besides that, the models are identical to the previously defined, fully specified regression models. Figure 11 depicts the point estimations and the 0.95 confidence intervals of the coefficients of the corruption proxies. Barely any statistically significant effects could be identified, with one exception. In the fixed effects model, Single Bidding has a significant positive effect on the share of patent applications belonging to the class *Other Fields*, which consists of the fields *Furniture and Games*, *Other Consumer Goods*, and *Civil Engineering*. These fields are associated with comparably low levels of technological complexity.

Figure 11: Effect of Corruption on Technological Composition



8 Discussion

To begin the discussion, I want to recap the main initial research question. Does corruption have a significant effect on innovation performance, and if so, is this effect positive or negative? At least for European regions in recent years, I found clear evidence supporting the argument that corruption sands the wheels of innovation. To extend the existing literature and fill prior research gaps, I did not only consider the commonly used number of patent applications as a measurement of innovation performance but also the respective quality of patents, as well as the technological complexity and technological composition of regions. Furthermore, I decided to apply two different model types, namely the between and within estimators. By doing so, I was able to isolate the two distinct dimensions of variation in the data, namely the variation between regions and the variation within regions over time. This allows for a more comprehensive picture while the interpretability stays intact, which can be seen as a clear advantage over limiting the analysis to either model type or applying an estimator that mixes between and within region effects [38].

Unsurprisingly, I found evidence that in Europe, corruption negatively affects innovation performance in terms of the number of patent applications. This aligns with the findings of Rodríguez-Pose and Di Cataldo (2015) [52]. However, my final research approach differed quite substantially, which shed light on some previously unconsidered issues. An especially interesting observation was that in the between estimation, Single Bidding does not have a statistically significant effect on the number of patent applications per capita, but the EQI does, while in the within estimation, this is reversed. Of course, this fact alone does not allow for any causal conclusions. However, one possible interpretation suggests itself. One of the biggest advantages of the Single Bidding indicator is that it is not based on perceptions and therefore does not suffer from stickiness or the problem that regions that performed economically well in the past are generally perceived as less corrupt [24]. On the other hand, the comparability between countries might be somewhat limited because even though the indicator is purely based on tendering contracts that appear in the TED database and should therefore be comparable, there is large heterogeneity between countries in terms of data quality and the overall number of contracts appearing in the TED [23]. The exceptionally low share of Single Bidding contracts in Greece, for example, indicates very high control of corruption, which does not align with other established indicators [14]. On the other hand, the perception based EQI may be less suitable for time-series analyses because of the persistence of perceptions. Future research aimed at analysing these hypotheses would be needed for validation and to support an informed, context-dependent choice of an appropriate indicator.

Besides confirming the negative effect of corruption on the number of patent applications, I could also find strong evidence that corruption has a negative effect on the quality of patents. This finding is consistent in the between and within estimation and broadly robust to the inclusion of the number of patents per capita as a control variable. This means that corruption affects patent quality directly and not only in a probabilistic manner via the number of patent applications. Another interesting observation is that it makes a big difference in patent quality and technological complexity whether the overall average or the average of the highest quality and highest complexity patents is considered. Depending on the context and research question, both approaches can be valid, with the latter being better suited to assessing a region's innovative capacity.

A further contribution of this thesis is that it establishes a relationship that was not analysed in the literature before, namely the effect of corruption on the technological complexity that a region can achieve. To do so, I used a recently developed measurement for technological complexity, namely the structural diversity of a technology's components [10]. While the evidence is not as robust as the measured effect on the quantity and quality of patents, the fixed effect models do suggest a negative relationship. Similar to the quality of patents, but to a greater extent, the difference between the overall average technological complexity and the average technological complexity in the 10% most complex patents is substantial. A counterintuitive observation is the strong negative effect of population density and the positive effect of the primary sector share within regions. Further research would be needed to assess this relationship in detail, but it seems that the urbanization of regions might lead to a decrease in technological complexity.

Given that corruption is associated with more complex innovation activities, the question arises of whether the level of corruption also affects the technological composition of a region. A potential hypothesis would be that more corrupt regions tend to engage in the research of less complex technologies because the more complex a technology is, the more actors would be involved, which gives more room for rent extraction and obstructive behaviour.

Overall, I did not find much supporting evidence for this claim, with one exception. In the fixed effects model, an increase in Single Bidding is associated with an increase in the share of patent applications belonging to the *Other Fields* class, which consists purely of low complexity technologies. Given that this result is not consistent in the between estimation and the within estimation with the EQI as corruption proxy, the robustness is severely limited, leaving us with a somewhat interesting result, but one that does not allow for causal inference.

The results clearly show that corruption negatively affects various facets of innovation performance in European regions. While this answers the main questions of this thesis, the question remains through which specific mechanisms and channels this relationship occurs. Coming back to the concept of innovation effectiveness and efficiency, there are two main channels through which the innovation performance of a region can be affected. The first is a region's capacity to attract and retain innovation generators [9]. Prior studies have shown that corruption is one of the main factors driving brain drain[4] [19, 49]. This is especially true for Central and Eastern European countries. Iacob (2018), for example, identifies corruption as the main push factor of Romanian skilled migration, even more relevant than salary levels, the healthcare system, professional opportunities, the educational system, and the quality of life [33]. With the east expansion of the EU and the accompanying opening of borders and free movement of labour, this trend has been accelerated significantly within Europe, with increasing brain drain from Central and Eastern to Western Europe [35]. Especially for innovators, the corruption induced reasons to leave a region and move to one with higher control of corruption are numerous. Rent-extraction, lacking protection of intellectual property, unreliable governmental institutions and a generally low level of trust amongst agents are all strong disincentives to reside and innovate in a corrupt region. Not only do these factors deter innovation generators from settling and remaining in a region, but they also impede cooperation, lead to less efficient processes of existing innovators and disincentivize companies and individuals to invest in innovation and other complex economic activities [5]. In such a frustrating environment for innovation, high potentials may even conclude that participating in corrupt rent extraction is more attractive and profitable than pursuing a career in research, leading to a self-reinforcing deterioration of the situation. Although all of these mechanisms are quite intuitive and based on the literature, I believe that there is much room for future research to empirically examine the specific mechanisms through which corruption impedes innovation performance to provide an evidence-based foundation for effective anti-corruption policy.

As for the implications of these findings for policymaking, the results mainly contribute to the existing strain in the literature that argues for the importance of anti-corruption measures in creating an ideal environment for innovation [52, 54]. While R&D expenditure is certainly an important driver of innovation, it is far from being the only one, as was shown in this thesis once again. To achieve the goal set by the European Commission to bring

[4]the emigration of highly educated citizens

Europe back to the forefront of innovation, a holistic approach will be needed that recognizes the role of corruption [18].

Of course, the results and conclusions derived in this study cannot be interpreted as depicting the absolute and final truth. Limitations arise mainly due to the data. First of all, the quantity of available data leaves room for improvement. I had to restrict my analysis to the years 2011 to 2017, which is not a long time period with regard to corruption and innovation performance, and even within this time period, there was substantial incompleteness. Missing data, as well as the harmonization of data, were issues that had to be dealt with, and even though the data processing was done with great care, a certain level of dilution of information cannot be avoided when imputation techniques are applied. Furthermore, there is no way to directly measure corruption and innovation performance. Hence, being tied to the use of proxies, the validity of the results may suffer. Both used corruption proxies have their own deficiencies. While Single Bidding is a somewhat indirect and noisy proxy for corruption, the EQI is only available for four sample years and has the same problems as other perception-based indicators, such as stickiness of perceptions. Nonetheless, the use of two separate proxies based on very different concepts helps increase the study's overall validity because even if neither of the proxies measures corruption perfectly, the detection of relatively consistent effects increases confidence that the results are meaningful.

Even in light of the aforementioned limitations, my findings add to a solid body of existing literature and extend previous studies in some key dimensions. Especially in the field of social sciences, one can never realistically claim to have found the truth, but the aim is to add convincing puzzle pieces to the whole picture to at least approximate the truth and come closer to it with every additional piece of literature.

9 Conclusion

Despite significant efforts and investments, the EU keeps falling behind in the global landscape of innovation. While many reasons for this are well-known, one aspect does not get broad attention, and that is the role of corruption. Different corruption proxies signal that, on average, European regions have become significantly more corrupt in the past decade. This is alarming in various ways, one of which is the negative effect of corruption on innovation performance. The results presented in this thesis have shown that corruption, at least in Europe, significantly sands the wheels of innovation. Corrupt regions produce fewer patents per capita, and the quality of patents is inferior. Furthermore, high levels of corruption are associated with lower technological complexity, which can lead to slower economic growth [45].

Some clear implications for policymaking can be derived. If Europe wants to regain its position as a key player in innovation, the current decline in control of corruption cannot be ignored and needs to be stopped. To reach this ambitious goal, we must acknowledge the immense heterogeneity in Europe that does not only exist between countries but also between sub-national regions and markets. Depending on the structure of the market, a differentiated approach to anti-corruption policy might be needed to maximise its effectiveness. In countries with highly centralized procurement markets, for example, it might be problematic when the central government is fully in charge of corruption control [57]. Therefore, regional decision-makers are also called upon to contribute to a less corrupt and more innovative Europe. Time will tell whether the EU will metaphorically succeed in rising like a phoenix from the ashes of the Covid-19 pandemic and regain its position as an innovation leader. A key determining factor will be whether policymakers acknowledge the importance of rigorous anti-corruption measures in creating an environment in which innovation can thrive and develop unhindered.

References

[1] Opentender. https://opentender.eu/start. Accessed: 2020-01-11.

[2] *Statistical regions in the European Union and partner countries : NUTS and statistical regions 2021 : 2020 edition*. Publications Office of the European Union, Luxembourg, 2020.

[3] Christopher H. Achen. Why lagged dependent variables can suppress the explanatory power of other independent variables, 7 2000. Presented at the Annual Meeting of the Political Methodology, Los Angeles.

[4] Zoltan Acs, Luc Anselin, and Attila Varga. Patents and innovation counts as measures of regional production of new knowledge. *Research Policy*, 31:1069–1085, 02 2002.

[5] Sergey Anokhin and William S. Schulze. Entrepreneurship, innovation, and corruption. *Journal of Business Venturing*, 24(5):465–476, 2009. Special Issue Ethics and Entrepreneurship.

[6] Daniel Antal. *regions: Processing Regional Statistics*, 2020. R package version 0.1.5.

[7] Andrew Bell, Malcolm Fairbrother, and Kelvyn Jones. Fixed and random effects models: making an informed choice. *Quality and Quantity*, 53:1051–1074, 03 2019.

[8] Andrew Bell and Kelvyn Jones. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. 3:133–153, 01 2015.

[9] Thomas Brenner and Tom Brökel. Methodological issues in measuring innovation performance of spatial units. *Philipps University Marburg, Department of Geography, Working Papers on Innovation and Space*, 18, 01 2009.

[10] Tom Brökel. Using structural diversity to measure the complexity of technologies. *PLOS ONE*, 14:e0216856, 05 2019.

[11] Tom Brökel. Updated values of technological complexity. https://www.tombroekel.de/updated-values-of-technological-complexity/ , 05 2020. Accessed: 2021-02-26.

[12] Tom Brökel, Nicky Rogge, and Thomas Brenner. The innovation efficiency of german regions – a shared-input dea approach. *Review of Regional Research*, 38, 03 2017.

[13] Nicholas Charron, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Aksel Sundström, Natalia Alvarado Pachon, and Cem Mert Dalli. The quality of government eu regional dataset, 2020.

[14] Nicholas Charron, Lewis Dijkstra, and Victor Lapuente. Regional governance matters: Quality of government within european union member states. *Regional Studies*, 48(1):68–90, 2014.

[15] Nicholas Charron, Victor Lapuente, and Paola Annoni. Measuring quality of government in eu regions across space and time. *Papers in Regional Science*, 98, 02 2019.

[16] Nicholas Charron, Victor Lapuente, and Monika Bauhr. Sub-national quality of government in eu member states: Presenting the 2021 european quality of government index and its relationship with covid-19 indicators. *QoG Working Paper Series*, 04 2021.

[17] European Commission. Horizon 2020 - the framework programme for research and innovation. Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 11 2011.

[18] European Commission. *Horizon 2020 in brief*. Publications Office of the European Union, 2014.

[19] Arusha Cooray and Friedrich Schneider. Does corruption promote emigration? an empirical examination. *Journal of Population Economics*, 29, 08 2015.

[20] Katrin Cremers, Dietmar Harhoff, Frederic Scherer, and Katrin Vopel. Citations, family size, opposition and the value of patent rights. *Research Policy*, 32:1343–1363, 09 2003.

[21] Francesco Decarolis and Cristina Giorgiantonio. Corruption red flags in public procurement: New evidence from italian calls for tenders. *SSRN Electronic Journal*, 02 2020.

[22] Charles Edquist and Jon Mikel Zabala-Iturriagagoitia. The innovation union scoreboard is flawed: The case of sweden - not being the innovation leader of the EU. 04 2015.

[23] Mihály Fazekas. *Assessing the Quality of Government at the Regional Level Using Public Procurement Data*. Publications Office of the European Union, 2017.

[24] Mihály Fazekas, Tóth János, and Lawrence King. An objective corruption risk index using public procurement data. *European Journal on Criminal Policy and Research*, 22, 09 2016.

[25] Mihály Fazekas and Gábor Kocsis. Uncovering high-level corruption: Cross-national objective corruption risk indicators using public procurement data. *British Journal of Political Science*, 08 2017.

[26] Maryann Feldman. The significance of innovation. 08 2004.

[27] European Commission's Directorate General for Regional and Urban Policy. Ardeco database. https://knowledge4policy.ec.europa.eu/territorial/ardeco-database_en. Accessed: 2020-03-08.

[28] Per Fredriksson, Richard Damania, and Muthukumara Mani. The persistence of corruption and regulatory compliance failures: Theory and evidence. *Public Choice*, 121:363–390, 02 2004.

[29] Iftekhar Hasan and Christopher Tucci. The innovation-economic growth nexus: Global evidence. *Research Policy*, 39:1264–1276, 12 2010.

[30] Hugo Hollanders and Adriana Cruysen. Rethinking the european innovation scoreboard: a new methodology for 2008-2010. 10 2008.

[31] Hugo Hollanders and Nordine Es-Sadki. *European Innovation Scoreboard 2017*. Publications Office of the European Union, 2017.

[32] Hugo Hollanders, Nordine Es-Sadki, Iris Merkelbach, and Khalilova Aishe. *European Innovation Scoreboard 2020*. Publications Office of the European Union, 2020.

[33] Raluca Iacob. Brain drain phenomenon in romania: What comes in line after corruption? *Romanian Journal of Communication and Public Relations*, 20:53, 07 2018.

[34] Transparency International. What is corruption? https://www.transparency.org/en/what-is-corruption. Accessed: 2020-12-23.

[35] Matthew Jelavic. The brain drain: Implications for regional economic integration in the expanding european union. In Bernard Chapelet and Michel Le Berre, editors, *Producing New Knowledge on Innovation Management*, pages 99 – 111. Presses Universitaires de Grenoble, 01 2012.

[36] Fatma Nur Karaman. Corruption and innovation: The case of eeca countries. 12 2017.

[37] Nathan Kelly and Luke Keele. Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis*, 14, 01 2004.

[38] Jonathan Kropko and Robert Kubinec. Interpretation and identification of within-unit and cross-sectional variation in panel data models. *PLOS ONE*, 04 2020.

[39] Leo Lahti, Janne Huovari, Markus Kainu, and Przemyslaw Biecek. eurostat r package, 2017. Version 3.6.84.

[40] Johann Lambsdorff. Consequences and causes of corruption - what do we know from a cross-section of countries? *Diskussionbeitrag Nr V-3405*, pages 3–52, 01 2006.

[41] Johann Lambsdorff and Gunther Schulze. What can we know about corruption? A very short history of corruption research and a list of what we should aim for. 235:100–114, 04 2015.

[42] Jean Lanjouw and Mark Schankerman. Patent quality and research productivity: Measuring innovation with multiple indicators. *Economic Journal*, 114:441–465, 02 2004.

[43] Debora Malito. Measuring corruption indicators and indices. *SSRN Electronic Journal*, 01 2014.

[44] Mara Mendes and Mihály Fazekas. Towards more transparent and efficient contracting - public procurement in the european union. https://opentender.eu/blog/assets/downloads/digiwhist_recommendations.pdf. Accessed: 2020-01-11.

[45] Lars Mewes and Tom Broekel. Technological complexity and economic growth of regions. *Research Policy*, 11 2020.

[46] Alex Opoku, Vian Ahmed, and Julius Akotia. Choosing an appropriate research methodology and method. *Research Methodology in the Built Environment: A Selection of Case Studies*, pages 32–49, 03 2016.

[47] Peter Phillips and Donggyu Sul. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *http://www.business.auckland.ac.nz/comwebContent/docs/Economics/WP252.pdf*, 09 2003.

[48] Marilen Pirtea, Gabriela Sipos, and Alin Ionescu. Does corruption affects business innovation? Insights from emerging countries. *Journal of Business Economics and Management*, 20:715–733, 06 2019.

[49] Marie Poprawe. On the relationship between corruption and migration: empirical evidence from a gravity model of migration. *Public Choice*, 163, 06 2015.

[50] Rudra Pradhan, Mak Arvin, John Hall, and Mahendhiran Nair. Innovation, financial development and economic growth in eurozone countries. *Applied Economics Letters*, 23:1–4, 02 2016.

[51] Muhammad Faraz Riaz and Uwe Cantner. Revisiting the relationship between corruption and innovation in developing and emerging economies. *Crime, Law and Social Change*, 73, 05 2020.

[52] Andrés Rodríguez-Pose and Marco Di Cataldo. Quality of government and innovative performance in the regions of europe. *Journal of Economic Geography*, 15, 07 2015.

[53] Mariagrazia Squicciarini, Hélène Dernis, and Chiara Criscuolo. Measuring patent quality: Indicators of technological and economic value. *OECD Science, Technology and Industry Working Papers*, 03 2013.

[54] Gindo Tampubolon. Corruption and innovation in european regions: building on the shoulders of giants. http://blog.gdi.manchester.ac.uk/corruption-and-innovation-in-european-regions-building-on-the-shoulders-of-giants/, 7 2018. Accessed: 2020-12-22.

[55] Nicolas van Zeebroeck. The puzzle of patent value indicators. *Economics of Innovation and New Technology*, 20, 09 2007.

[56] Nicolas van Zeebroeck and Bruno van Pottelsberghe. The vulnerability of patent value determinants. *Economics of Innovation and New Technology*, 20:283–308, 04 2011.

[57] Johannes Wachs, Mihály Fazekas, and János Kertész. Corruption risk in contracting markets: a network science perspective. *International Journal of Data Science and Analytics*, 01 2020.

[58] Hua Xu, Sock Hwan Lee, and Tae Eom. Introduction to panel data analysis. In Gerald J. Miller and Yang Kaifeng, editors, *Handbook of Research Methods in Public Administration*, pages 571–590. CRC Press, 11 2007.

[59] Robert Yaffee. A primer for panel data analysis. *Connect*, 01 2003.

[60] Štefan Šumah. Corruption, causes and consequences. In Vito Bobek, editor, *Trade and Global Market*, chapter 5. IntechOpen, Rijeka, 2 2018.

A Summary Statistics

Table 13: Summary Statistics - Patent Variables

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
N Electrical Eng	51.40	112.96	0.00	2.50	42.04	947.56
N Instruments	34.04	73.14	0.00	2.08	33.20	1,033.42
N Chemistry	48.25	87.76	0.00	4.30	52.80	741.00
N Mechanical Eng	70.04	126.74	0.00	5.70	78.18	1,284.25
N Other Sector	21.44	36.18	0.00	2.00	24.09	251.46
N No Sector	0.18	1.17	0.00	0.00	0.00	38.00
N Total Patents	225.35	394.94	0.17	20.21	263.30	3,477.22
Mean Quality	0.28	0.06	0.06	0.25	0.31	0.59
Max Quality	0.57	0.16	0.06	0.47	0.69	0.93
Top 10% Quality	0.47	0.11	0.06	0.42	0.53	0.81
90^{th} Pctl Quality	0.41	0.09	0.06	0.37	0.46	0.81
Mean Complex	11.60	0.36	10.04	11.39	11.82	13.29
Max Complex	13.23	0.59	10.10	12.97	13.75	13.95
Top 10% Complex	12.73	0.38	10.10	12.59	12.95	13.95
90^{th} Pctl Complex	12.54	0.37	10.10	12.36	12.75	13.76

Table 14: Summary Statistics - Corruption Proxies

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Single Bid	16.37	14.29	0.00	5.63	23.48	100.00
No Call	25.61	24.25	0.00	6.78	37.59	100.00
Non Open	7.12	10.67	0.00	0.28	8.92	100.00
Tax Haven	2.73	11.90	0.00	0.00	0.00	100.00
EQI	55.82	19.78	0.00	41.09	71.10	100.00
EQI Corrupt	56.96	19.89	0.00	39.77	71.52	100.00
EQI Impart.	59.45	18.05	0.00	45.54	73.19	100.00
EQI Quality	58.88	18.45	0.00	47.01	72.89	100.00

Table 15: Summary Statistics - Control Variables

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
GDP PPS	27,010.83	12,920.43	7,400	19,800	31,700	187,300
Unemployment	9.12	6.22	1.70	4.80	10.80	36.10
Pop. Density	480.47	1,266.60	3.30	76.15	320.35	11,357.10
Primary Sector	5.74	7.04	0.20	1.70	6.85	52.50
Tertiary Ed.	28.69	9.68	9.90	21.55	34.40	74.70
Hightech	3.44	1.87	0.60	2.10	4.25	11.00
R&D Spending	515.77	584.32	4.20	124.55	656.25	3,884.30

B Correlations and Maps of Control Variables

Figure 12: Correlation Matrix of Control Variables

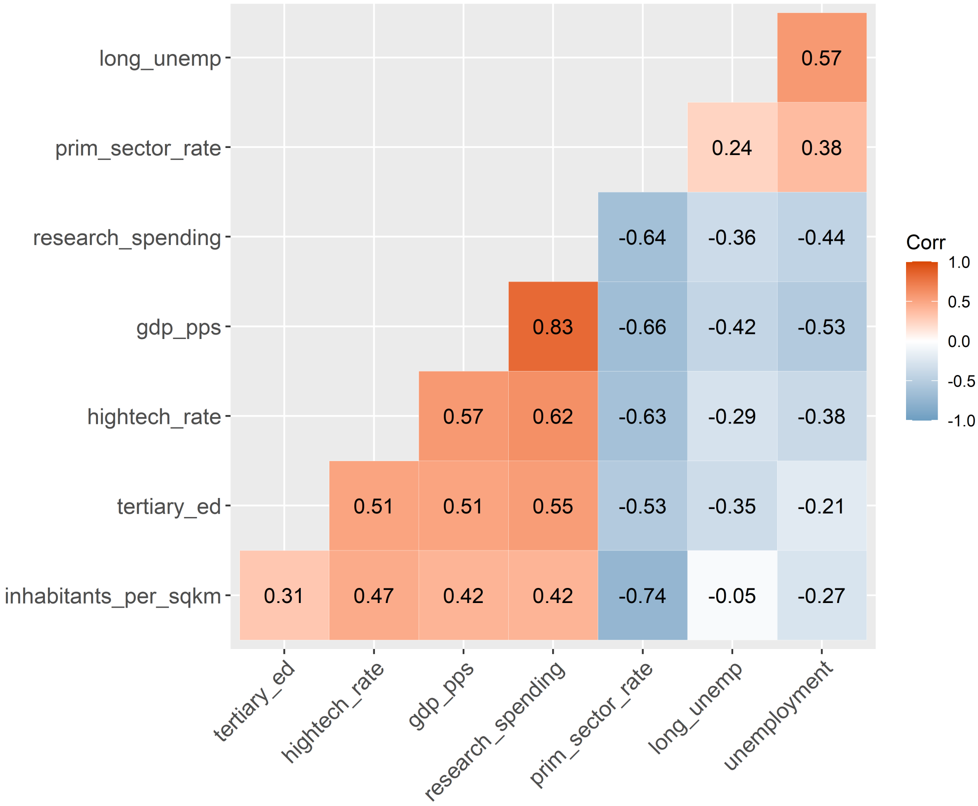


Figure 13: Map of GDP

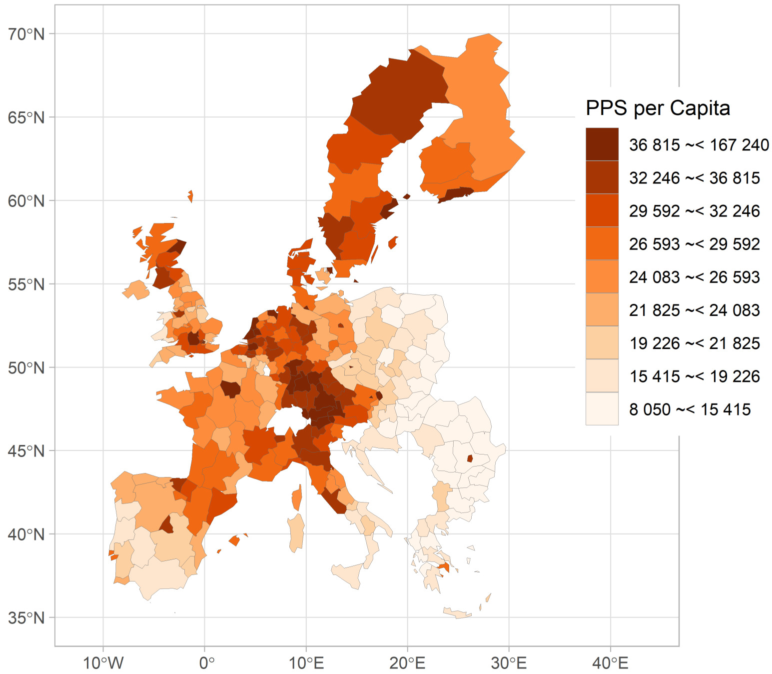


Figure 14: Map of Unemployment

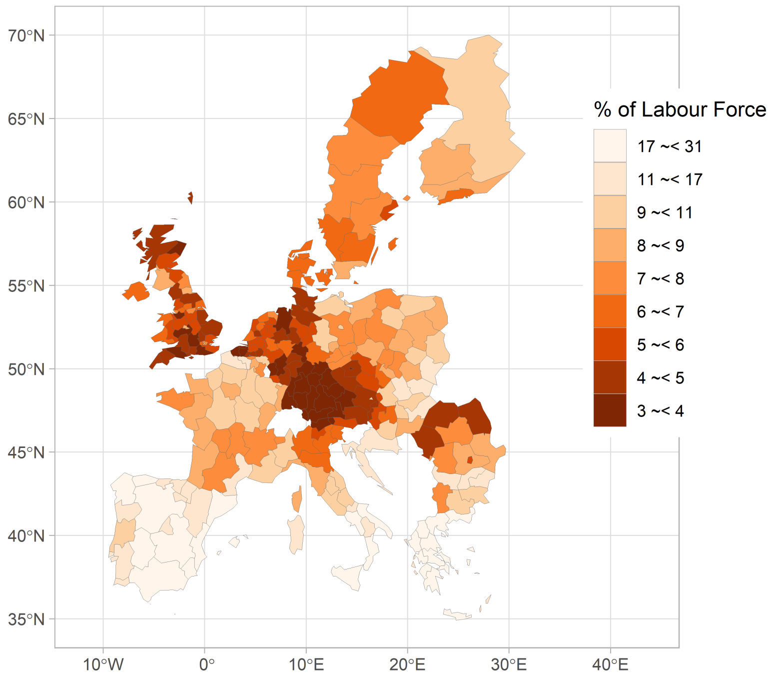


Figure 15: Map of Population Density

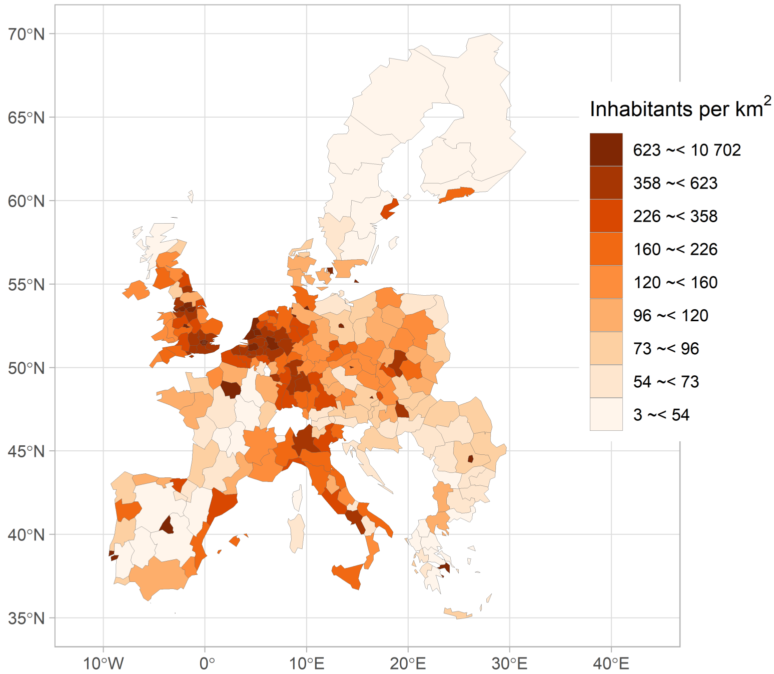


Figure 16: Map of Primary Sector Employment

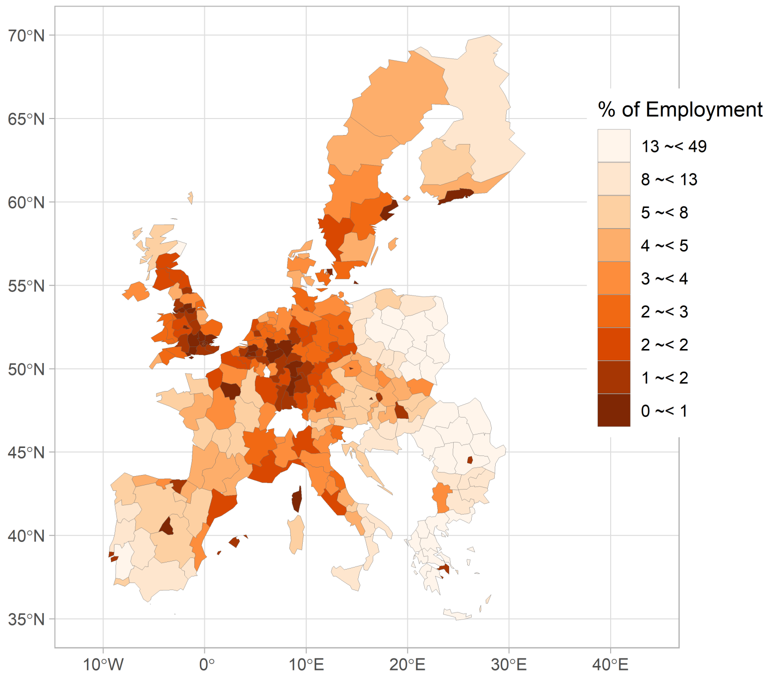


Figure 17: Map of Tertiary Education Rate

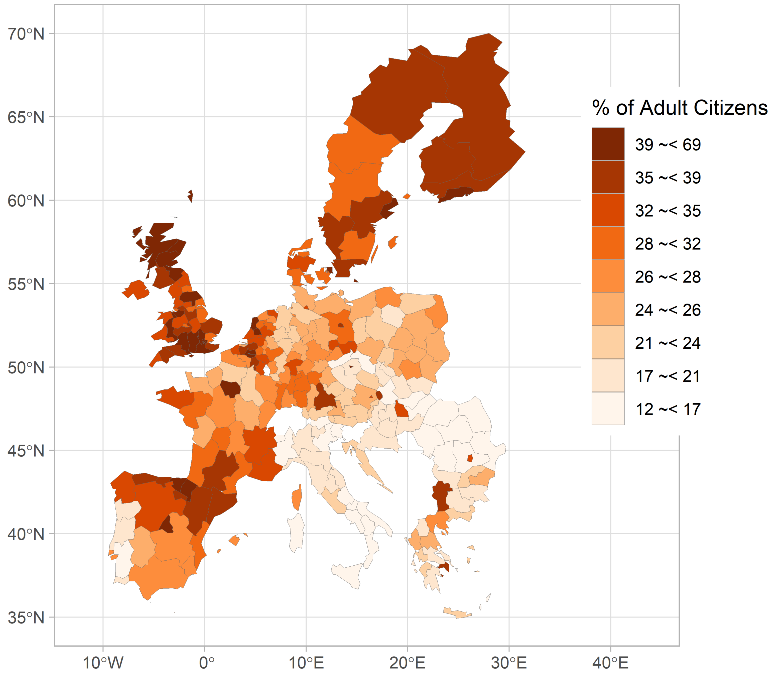


Figure 18: Map of Hightech Employment

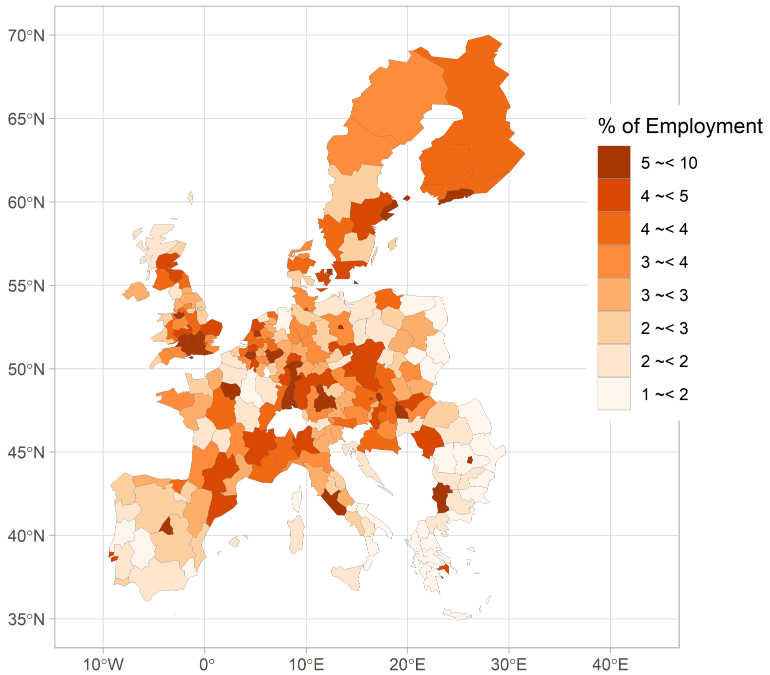


Figure 19: Map of R&D Expenditure