ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS URBAN, PORT & TRANSPORT ECONOMICS

The Underlying Relationship of FDI Determinants: A Study on European Metropolitan Clusters

Stijn Riemslag – 483845

Supervisor: M.A. Rosch

Second assessor: Z. Wang

Date final version: April 28, 2023

Abstract

This paper investigates the underlying relationship of characteristics of European economic clusters and to what extend they attract inward greenfield foreign direct investment (FDI). Three main factors are distinguished, based on the OLI framework of Dunning (2000); the Corruption Perceptions Index (CPI), the global interconnectedness and the knowledge sector specialization of the clusters' main activity. The characteristics are combined and compared with the effective average tax rate (EATR) of corporations and it is argued why these attract FDI, based on earlier literature and international business. The arguments are tested using negative binomial regressions on data of European metropolitan clusters over the period of 2003-2018. The results indicate a positive relationship between the number of greenfield investments in a cluster and both the regions' CPI score and specialization in the knowledge sector. The EATR is negatively related to the number of greenfield investments with a diminishing non-linear effect, due to a positive squared term. Evidence for a tradeoff effect is presented by marginal effects at different tax rates, showing the decreasing effect of corporate taxes and increasing effect of a component of global interconnectedness, at increasing corporate tax levels. Lastly, the results are linked to policy decisions regarding FDI determinants and the attraction of greenfield FDI towards European economic clusters.

Keywords – Foreign direct investment, clusters, tradeoff, location strategy

Contents

1 Introduction

Since foreign direct investment (FDI) increased significantly over the last few decades, it has become a crucial topic of discussion in the field of economics (UNCTAD, 2022). The increasing number of multinationals investing in foreign countries leads to an increase in cash flows worldwide. The receiving part of the investments, local economies, can benefit from these FDI flows as the economic growth in the region is positively influenced. The benefits of FDI for local economies are measurable in factors such as increased (technological) development, productivity, and knowledge spillovers, which are positive externalities of receiving foreign investments (Rasciute & Downward, 2017; OECD, 2019).

As a result, the local economies are in competition, trying to maximize their FDI attraction by implementing policies that provide a beneficial economic environment and competitive advantage for the investing multinational (MNE). However, the competition between FDI is not equally distributed and consists of different regions competing for different investment streams. This research investigates the tradeoff between the policies on different determinants of attracting FDI.

First, the regions are very different regarding their natural resources and accessibility. Second, to compete for FDI, a region must have reached a minimum level of development in education, technology, and infrastructure. In other words, FDI per se only provides direct growth opportunities if the receiving region has the necessary capacities to profit from the externalities of the investment activity (Lall & Narula, 2004; Hansen & Rand, 2006). Because the competition and characteristics for FDI can differ between the regions, different types of FDI streams are divided into multiple receivers, which leads to a spread of FDI streams throughout various territorial regions.

Although the advantages of natural resources are fixed, regions can influence their policy decisions regarding FDI to strengthen their competitive position. This can be done by focusing on multiple factors. A common way of attracting investments when competing over FDI is to specialize and focus on specific potential developments that complement a region's economic structure (Gordon, 1999). Porter (1998) describes how the clustering of economic activities in urban regions is profitable. It can lead to agglomeration economies, where benefits arise from firms and industries location close together in clusters. Beneficial factors are, for example, an enhanced reputation, knowledge spillover, and the ability to use a specialized labor pool. Furthermore, Porter (1998) finds evidence of positive spillovers between firms and patterns of clustering. Clusters are unusually productive, ensuring an increase in resource flows to the cluster and enabling the re-enforcing of its advantages. These advantages of agglomeration economies make clusters magnets for national and international talent and investments (Cook et al., 2004; Crozet et al., 2004). The arguments state that clusters can profit from potential agglomeration effects by bundling knowledge and resources to benefit from positive spillovers and increase their attractiveness regarding FDI investments.

Burger et al. (2013) show that (tax) policies can significantly affect territorial competitiveness regarding FDI. They describe how rule-based policies regarding the regulation of trade are, for a large part, generalized in Europe, leading to the increase in the importance of the level of corporate taxes (incentive-based policy). The authors highlight that low corporate taxes pose one of the most significant competitive threats to other European regions. This argument is also visible in recent developments in European countries, with two large multinationals (Shell and Unilever) leaving The Netherlands due to disagreements with the governments about the level of corporate taxes (Currie, 2017). Corporate tax rates are defined as an incentive-based policy because it is an immediate policy that could be implemented in a relatively short term. Regarding FDI, changing the corporate tax rates can be used as a potential way to increase the region's attractiveness. Existing literature presents a negative effect of the taxation rate on the probability of choosing a location in the receiving country (Barrios et al., 2012). Lawless et al. (2018) combine corporate taxation and the location choices of MNE by testing for a non-linear approach instead of a linear approach, which implies that changes in tax rates at higher levels have a smaller effect. They find a strong negative non-linear effect of taxation on the likelihood of a destination being chosen. Both findings suggest that the receiving countries or regions of FDI investments could differ significantly based on their corporate tax rates. Thus, the corporate tax rates are an essential policy of the European countries with the potential use to increase attractiveness. Hence, the rates differ between the countries and over time (de Mooij & Nicodème, 2008; OECD, 2022a).

It is also possible to increase the country or region's attractiveness with policies that take longer to change the competitive position regarding FDI effectively. Creating a favorable environment based on a political or economic nature can take years. Examples are improving the political environment or working on projects that can increase the accessibility of a cluster. Based on previous arguments, regions attracting FDI have multiple policy options to increase their competitive position regarding FDI. For faster incentive-based policies, the rates of corporate taxes can be adjusted. For a more stable and long-term increase of the FDI determinants, industries and firms can cluster in agglomeration economies, and multiple-year projects (e.g., infrastructure) can be executed. This creates a tradeoff in the policy decisions regarding FDI for regions and firms (Brülhart et al., 2012).

The presented literature regarding this subject has its limits. Multiple articles write about the significance of location choices of FDI determinants such as corporate taxes, infrastructure, and agglomeration economies (Lall & Narula, 2004; Burger et al., 2013). However, combining the aspects with a potential tradeoff effect for FDI determinants in regions remains underexposed. Furthermore, most of the investigations are focused on a national or regional dataset, with little attention to external validity (Brülhart et al., 2012). This research aims to look at the effect of the underlying relationship of the FDI determinants with the corporate taxes on the one side and the remaining FDI determinants on the other. As a result, this research complements the literature on territorial competition by investigating a tradeoff effect between the characteristics. To do so, the following research question is central:

How do agglomeration economies and corporate taxes interact in attracting Foreign Direct Investments?

In order to answer this question, insights from the existing literature will be followed and projected on firm and country-level data of the European Union. This will be done by analyzing greenfield investments from the European clusters over the period of 2003-2018 retrieved from the fDi Markets (2022) database. The dataset contains information on the complete process of the FDI stream from the company of origin to the city of destination. To analyze the territorial competition between the regions, a cluster classification will be made based on the European metropolitan regions (Eurostat, 2023a). In line with the aim of this research, greenfield data is used to collect the investments of new ventures in foreign countries to capture the sensitivity of MNEs towards the changing FDI determinants. With this information, the paper presents multiple (sub-)hypotheses to define and analyze the aim of the research.

First, attention will be brought to FDI and clusters. The European situation will be used to analyze and follow the OLI paradigm from the past literature of Dunning (2000). FDI determinants from this framework, the level of corruption, global interconnectedness, and specialization, will be tested in a count data model with multiple independent variables. The variables are selected because they interact with policies that take longer to change than the tax policies. The level of corruption is created in the political environment and is a long-term factor. The global interconnectedness represents the region's accessibility and is often only increased with long, multiple-year projects. The specialization measures a specific knowledge cluster in the region and enables us to identify agglomeration benefits. Secondly, the research directs attention toward the involvement of corporate taxes. Tax data on the effective average tax rates will be added to investigate the relationship between location choices and corporate taxes. This will be analyzed while considering the non-linear relationship of corporate taxes (Lawless et al., 2018). Furthermore, this article tests the importance of the FDI characteristics to visualize the tradeoff with corporate taxes.

The expected effects of the tradeoff in territorial FDI competition are twofold: (1) corporate taxes are expected to stand out significantly. As Burger et al. (2013) describe, the weight of corporate taxes is high in European countries, and the characteristic will therefore be more decisive. Furthermore, expectations align with the findings of Lawless et al. (2018), who describe a non-linear effect of taxation on the likelihood of a destination being chosen for the investment. (2) clustering of economic activities based on favorable FDI characteristics is beneficial for attracting greenfield investments. A similar result was found in past literature but will be used in the context of the tradeoff, implying that the non-linearity influences the importance of all other characteristics (Crozet et al., 2004; Cook et al., 2004). As a result: due to the non-linear effect of taxation, the corporate taxes in the tradeoff are expected to be more important in a situation of relatively low taxes. The remaining FDI characteristics are expected to grow in importance in the tradeoff when the corporate taxes increase. This requires research not only on different characteristics but suggests the need for investigating the tradeoff as a whole and the underlying relation of all the included parts.

The relevance of this paper can be found in the tradeoff of policy decisions regarding different FDI determinants. Multiple papers describe separate topics of agglomeration economies, FDI developments, or corporate taxes in relation to locational investment decisions. However, the combination of determinants (with the expected tradeoff effect) remains underexposed (Basile et al., 2008). Furthermore, this research gives an overview of European greenfield investments with up-to-date data. To address the various subjects, this research will continue with the theoretical background of the different topics related to FDI, cluster competition, and corporate taxes. This is followed by hypotheses that will contribute to answering the research question. The remainder of the paper is structured as follows. Sections 4 and 5 will discuss the data and methodology, respectively. Section 6 will discuss the results, and sections 7 and 8 will state the robustness checks and concluding remarks.

2 Literature Review

This research connects multiple streams in the literature. It aims to visualize the different FDI determinants and find the relation that the determinants have with each other within agglomeration economies. To do so, this section starts with the practice of agglomeration economies and the relationship with FDI. Hereafter, the focus will be shifted to the FDI determinants and their underlying relationship. The OLI paradigm will be used to select multiple determinants from the FDI framework and connect them to the literature streams (Dunning, 2000).

2.1 Agglomeration and FDI Attraction

The first part will investigate the agglomeration of European clusters and their relation with FDI attraction. Clusters generally consist of one or more European cities within the same country, sharing a variety of linked industries, (governmental) research institutions, and other entities important to competition (Porter, 1998). Therefore, a distinction can be made between the FDI determinants of the host regions, such as the business environment, governmental factors, and capabilities. Furthermore, Franca et al. (2021) highlight the clusters' role in improving regional resilience and their role in being a collaborative bridge between businesses and policymakers.

One of the functions of agglomeration in certain areas is to anticipate urban and economic complementarities. Furthermore, agglomeration economies include benefits when firms and people locate together in cities and industrial clusters (Glaeser, 2010). First, nearby clustered firms are more accessible, ensuring various possibilities and a favorable business environment. Second, clustering is associated with the potential for external economies of scale and spillovers. This explains how firms can produce at lower costs while benefitting from lower transaction costs, a wider range of opportunities for matching needs and capabilities, and the exchange and spillover of knowledge (Scott, 1992; Kloosterman & Lambregts, 2001). In addition, worldwide development gives broader access to technological capacities and decreases transportation costs for goods, people, and ideas. This suggests clusters become less important over time. However, a central paradox is described by Glaeser (2010), who finds that agglomerations remain remarkably significant, argued by investigating recent developments in agglomerations and revealing the (positive) trends in house prices, productivity, and population creation.

Regarding FDI flows, research finds that the attraction depends on changes in the landscape of the international business environment and that it supports international trade with the ability to create stable and long-lasting links between economies (UNCTAD, 2019, 2022). Therefore, FDI is a tool for providing growth (opportunities) in the regions and supports their international connections. Furthermore, the investment flows present direct incentives for the local economies on a large scale with a dynamic nature and indirect effects through positive externalities linked to the presence and investments of foreign MNEs. The positive externalities in agglomeration economies are found in the spillovers of technology and knowledge, which augment the existing stock of technological capabilities and knowledge of the host country (Bijsterbosch & Kolasa, 2010; Mehic et al., 2013). As a result, the specialization and agglomeration economies within clusters create a beneficial starting point for attracting investments based on specific FDI characteristics. Therefore, the clusters compete based on their capacities to profit from the externalities generated by the investment activity. However, due to the nature of the capacities (e.g., static capacities like natural resources), the investments spread streams throughout different territorial regions (Lall & Narula, 2004; OECD, 2019).

As stated before, an effective way of attracting investments when competing over FDI is to focus on specific potential developments that complement a region's economic structure. The interaction with agglomeration economies is found in the benefits of clusters regarding specific FDI determinants. With spillovers, scales of economies, and lower transaction costs, the regions are beneficial in competing for the attraction of FDI. The agglomeration economies can help regions capture the fullest potential of the growth opportunities that come with attracting FDI investments. Therefore, regions can stimulate agglomeration economies in certain locations with FDI determinants such as tax policies and infrastructure improvement. However, it is also influenced by external factors such as natural resources, knowledge spillovers, and economies of scale. These characteristics improve the stability of the European economy, supporting international trade and having the ability to create stable and long-lasting links between economies (Franca et al., 2021).

2.2 The Role of FDI Determinants

After stating the importance and interaction of agglomeration economies in attracting FDI, the focus will be shifted toward the determinants of the investments. The main determinants for the FDI flows of MNE activities are in the OLI paradigm, which is the preeminent theoretical framework for FDI decisions. In this framework consisting of the Ownership, Location, and Internationalization sub paradigms, multiple FDI characteristics that are decisive for attracting FDI flows can be categorized (Dunning, 2000).

The ownership sub-paradigm asserts that greater competitive advantages of investing firms lead to a higher likeliness of engaging in or increasing their foreign production (Dunning, 2000; Blanton & Blanton, 2007). In other words, for the receiving firm in the cluster of the FDI stream, providing these competitive advantages enlarges the likeliness of attracting FDI. Ownership factors determining the possibility of presenting competitive advantages are usually intangible and can be found in multiple characteristics that include proprietary information and various company ownership rights. Combining earlier research about the OLI Paradigm from Dunning (1973, 1982) presents arguments for ownership characteristics such as size, growth, and environment as decisive factors. Moreover, the focus is on the value-adding level of the characteristic, where intangible assets of firms in management skills and the cultural, legal, and institutional environment are highlighted (Pitelis, 2009; Rugman, 2010; Vijayakumar et al., 2010). Within this environment, Abdella et al. (2018) find that a higher level of political stability positively influences the attraction of FDI in a region. On the other hand, corruption is negatively correlated with FDI as business ethics can differ between people or because business ethics differ in their minds (Gopinath, 2008). Therefore, the effect of a region's political situation and environment can be twofold, and it remains an important determinant for attracting FDI.

In the second sub-paradigm, location, advantages are gained by going abroad, such as savings in transport costs, access to natural resources, getting around trade barriers, or the proximity to a large market. The advantages benefit MNEs since they are generally attracted to locations where the cost of communication between different regions is low. Therefore, the degree of interconnectedness increases, and the transport and transaction costs decrease, causing a positive effect on the attraction of FDI (McCann & Shefer, 2004; Blanton & Blanton, 2007; Bel & Fageda, 2008). In Europe, a large intermodal network is the transportation model for a person or goods from its origin to its destination by a sequence of multiple transportation modes. The transportation between clusters can be arranged by air, rail, road, or water, and (most) clusters are linked through all four aspects. In this European infrastructure, the global interconnectedness is very high through multiple important transport nodes on a larger scale and a very well-connected infrastructure on the European scale, such as the Port of Rotterdam and the connection to its hinterland (Port of Rotterdam, 2021). Therefore, the global interconnectedness of the European clusters is established through high accessibility in their intermodal infrastructure network.

The internationalization part of the OLI framework describes how MNEs can gain advantages and be more cost-effective when producing in-house while operating in a different market (Blanton & Blanton, 2007). Dunning (2000) explains that (some) firms, given a set of ownership and locational specific advantages, prefer to own their foreign value added rather than use their ownership advantages to independently owned foreign firms. If firms decide to outsource, creating partnerships with local firms is essential for obtaining advantages of outsourcing in market knowledge or more skilled employees. However, outsourcing only makes sense if the local firms' quality standards are similar and costs are lower. Therefore, the internationalization part of the OLI framework is mainly dependent on the ownership characteristics of the investing firm and the characteristics of the receiving market. Regarding these market characteristics, the incentives are in the advantages of the receiving region. Hence, a skilled and specialized labor pool that works more efficiently can reduce costs or increase production while keeping the same quality. Burger et al. (2013) further describe the importance of specialization by highlighting it as one of the key advantages of cluster regions. Furthermore, they distinguish between different sector specializations as some are preferred over others. The most attractive sectors, such as software and financial and business services, are knowledge-intensive, while more labor-intensive sectors are neglected.

The arguments explain the OLI paradigm using FDI determinants that influence national and regional regions. The sub-paradigms of ownership and internationalization are not explicitly defined but can be adapted by agglomeration economies. Clusters are, for example, known for their economies of scale, which can be argued as a part of the ownership sub-paradigm while creating competitive advantages. Dunning's (2000) paradigm also interacts with agglomeration economies for the locational and internationalization characteristics. With lower costs of communication (location) and, for example, the possibility to use a specialized labor pool (internationalization), clusters can exploit the possibilities of the paradigm.

2.3 Involvement of Corporate Taxation Policies

Despite the removal of many European trade barriers by making agreements, such as rules on matters in sustainable development or intellectual property, certain governmental policies remain essential in the environment of foreign investments (European Union, 2023). Following the research from Burger et al. (2013), taxes are crucial in current policies as low corporate taxes are identified as one of the biggest threats to other European clusters. One of the reasons is that tax policies are incentive-based, which is an immediate process and can be changed anytime.

Due to different governmental policies, corporate taxes are generally non-linear and can differ between countries, regions, and companies (Devereux & Griffith, 2002). Corporate taxes can also be lowered individually by deductions such as government subsidies or finding tax loopholes. This creates a difference between the statutory corporate income tax rates (policy rates) as levied by the governments and the effective average tax rates (EATR) that corporations pay. The effective tax rate in previous literature consists mainly of a total average based on the average of the effective marginal tax rates (Lawless et al., 2014, 2018; OECD, 2021, 2022b). To determine the investment amount and weight on corporate taxes, companies find the point where the expected pre-tax rate of return is equal to the cost of capital (in a hypothetical situation where other determinants do not play a role). To capture these effects, marginal taxes are used as an effective measure. Furthermore, research has shown that the tax deduction possibilities causing the differences in effective tax rates vary between countries. This causes the EATR to be generally lower than the policy rates and increases the differences between regions (Janský, 2019; Devereux & Griffith, 2002, 2003).

In Europe, the differences between statutory corporate rates and the EATR have increased the focus on matching taxation policies. Caused of the tax-avoiding practices of the MNEs, the European Union implemented multiple measures to harmonize taxation across its member states, such as the Common Consolidated Corporate Tax Base (CCCTB) and the Anti-Tax Avoidance Directive (ATAD). Both measures can potentially reduce tax avoidance and increase the revenue for the European member states. However, resistance from disadvantaged member states and a limited impact of the reforms caused by new company structures that avoid the rules lower the efficiency of such measures (European Commission, 2016, 2019). Therefore, the tax differences in Europe are still visible in both policy rates, generally around 20-25%, and in the EATR, which can differ even more (Janský, 2019; Statista, 2021; European Commission, 2022).

With fluctuations in policy rates and the EATR of regions, the MNEs have a significant factor to consider in their FDI location choices. Categorizing corporate taxes as an FDI determinant implies that it affects the FDI flows in a certain way. Barrios et al. (2012) find a significant and negative effect on the EATR of the host country and the probability of that location being chosen for the FDI flow. Combining the differences in tax rates, policies, and the negative effect on investments highlights the importance and influence of the tax presence in the FDI location decision of MNEs.

2.4 The Tradeoff in FDI Determinants

Earlier literature has shown the effects of various FDI determinants on the territorial competition between European clusters. Multiple characteristics present various effects on the attraction of FDI, and the mix of decisive determinants varies for each cluster based on market position, supply, and demand. As a result, the FDI characteristics are influenced by agglomeration economies positively and negatively, depending on the set of determinants, such as the influence of corporate taxes, which reduces the attraction of certain regions (Brülhart et al., 2012; Burger et al., 2013). However, other determinants, such as infrastructure, market size, and the political situation, also impact FDI attraction (Goodspeed et al., 2007). This creates a tradeoff scenario for the regions receiving investments in their FDI-related policy. Based on their specific mix of determinants, the regions can efficiently plan their policy measures.

Following these arguments, a division in FDI determinants can be made based on their policy effect and implementation speed. The effect of the policies is relative. Casi and Resmini (2014) describe how the determinants of FDI are interlinked and estimate the effects based on the relative performance of the region and its country. In other words, since each region has its own specific set of FDI determinants, they differ in importance and effect on the attraction of investments. Although corporate taxes pose one of the largest threats to territorial competition between regions, this is assumed to be the result of the speed at which new tax policies can be implemented (Wijeweera et al., 2007; Burger et al., 2013). Other FDI determinants, such as infrastructure policies, often consist of long-term projects that can take multiple years to finalize (Farhadi, 2015).

Based on the different perspectives on the relative importance of FDI determinants for regions, a tradeoff effect is assumed on the underlying relationships. This tradeoff of policies regarding the determinants depends on the implementation speed and effect and therefore influences the attractiveness of certain clusters in multiple ways.

3 Economic Theory

This section presents the theoretical argument developed in multiple hypotheses that will contribute to answering the research question. The hypotheses are based on findings from earlier literature on the geographical aspect of firms' location choices regarding FDI. The research considers multiple FDI determinants following the OLI framework and challenges them in a potential tradeoff with corporate taxes.

3.1 OLI Determinants

Political situation

The cultural, legal, and institutional environment is highlighted in the ownership sub-paradigm by Abdella et al. (2018). As stated before, the effect of political instability on the attraction of FDI can be negative, as described in the correlations between corruption and FDI (Gopinath, 2008). Moreover, corruption can negatively affect the likelihood of FDI taking place at the individual firm level, but it can also affect the firm choices at the aggregate level (Barassi & Zhou, 2012). Regarding policy implications, studies show that corruption affects the FDI location decisions of MNEs differently. It makes the local bureaucracy less transparent and hence adds to the cost of doing business. A counterargument of corruption is that it affects the decision of a local joint venture (JV) partner. Such a foreign partner increases in value by being present in a 'corrupt' foreign that is otherwise not open to foreign activities. However, this effect is neglectable because a local JV in a corrupt country increases the risk of misuse. Therefore, research indicates that a higher level of corruption would deter FDI. However, the effect is less when not only greenfield investments are considered and once a country is selected as a host country (Hakkala et al., 2008; Javorcik & Wei, 2009). Combining the mentioned arguments for the correlation between corruption and FDI, the following hypothesis can be stated:

H1a: Corruption in the host region deters the attraction of greenfield FDI for European economic clusters.

Global interconnectedness

The locational advantages of agglomeration economies positively affect the attraction of FDI in a region. This is supported by literature finding a positive effect of interconnectedness on the attraction of FDI caused by decreased costs of communication, transport, and transactions (McCann & Shefer, 2004; Blanton & Blanton, 2007; Bel & Fageda, 2008). Although most European clusters have multiple characteristics of global interconnectedness, there is no equal distribution across the regions. Multiple large transport hubs and parts of the (high-speed) infrastructure network are therefore (not) located in specific parts of the continent. However, the unequal distribution is in proportion to population numbers and density. Hence, a counterargument might be that the differences in accessibility result from urbanization effects. However, even if this was the case, global interconnectedness is a driving factor for trade with many large European ports and airports, such as the Port of Rotterdam and Paris Charles de Gaulle Airport (Eurostat, 2022b).

The intermodal transport options, such as transportation by air or rail, are primarily between densely populated central regions in Europe, where major cities are at relatively smaller distances apart (Spiekermann & Wegener, 1996). This causes a skewed distribution of the interconnectedness and locational advantages between clusters. Combining the literature on the effect of interconnectedness on the attraction of FDI and the skewed accessibility of the European clusters, the following hypothesis is formulated:

H1b: A high degree of global interconnectedness attracts inward greenfield FDI for European economic clusters.

Specialization

Specialization as an FDI determinant is one of the key advantages of agglomeration economies (Gordon, 1999). Burger et al. (2013) describe the importance of the type of specialization, highlighting the interest in investing MNEs in knowledge sectors and neglecting labor-intensive sectors. The European situation states that most cities and regions have activities in both sectors. However, the main and largest activities can be classified between one of the two specializations and are not equally distributed across Europe, such as the banking knowledge-intensive specialization for the London cluster (Cook et al., 2004).

Clusters target specific investments to attract FDI flows that complement their economic structure. This requires less effort, and establishments are better embedded in the economic structure, making it less likely for firms to reconsider their location and more likely to make additional investments at a later stage (Burger et al., 2013). The reason for the increased interest in knowledge-based sectors is the growth potential. With increasing technology developments and communication, the capacities are provided to profit from the externalities, creating a competitive advantage for the investing MNEs and stimulating the local economy of the host region (Lall & Narula, 2004; Hansen & Rand, 2006). Based on the MNEs' attraction towards the knowledge-intensive sectors and the differences between European regional clusters, the following hypothesis is formulated:

H1c: The attraction of inward greenfield FDI towards European clusters is stronger for knowledge-intensive sectors than for labor-intensive sectors.

3.2 Effective Average Tax Rates

Numerous studies show the importance of tax policies on the FDI location decisions of MNEs. Although continent-wide measurements stimulate the harmonization of taxation, differences in (mostly EATR) rates are still visible (Lawless et al., 2014; Janský, 2019). Furthermore, the EATR has a negative effect on the probability of a location being chosen for FDI (Barrios et al., 2012). Similar findings are presented by Davies et al. (2009), who conclude that tax policies affect the probability of investment in a certain region but not the volume invested by the FDI of firms.

Combining the literature on tax policies and FDI location decisions with the non-linear nature of the tax systems in European regions, a different approach is described by Lawless et al. (2018), who test for a non-linear response of firm location decisions to differences in tax rates. They find that accounting for potential non-linearity in the tax effects improves the model's performance for all the alternative tax rate measures. The measures show a significant negative effect of taxation on the probability of choosing a location but present a positive squared term. This indicates that the strength of the negative effect of taxes moderates as the rate of tax increases. In other words, for regions with lower corporate taxes, tax rate increases have greater relative effects on the probability of a location being chosen as the destination for the FDI flow. Lawless et al. (2018) also test for the different tax measures by distinguishing effects in the policy rate and the EATR. Following their findings, an increase in the policy rate reduces the conditional location probability, which is lower than for the EATR. This result confirms the finding of visible differences in European statutory corporate taxes and even larger differences in the EATRs (Janský, 2019).

The sensitivity of corporate taxes is further explained by the finding that the effect of tax policies differs between various countries and that the effect of the taxes is higher at a lower tax rate (Morrisset & Pirnia, 1999; Abdioğlu et al., 2016). This suggests a non-linear and negative relationship between the EATR and the probability of a chosen location. Combining this with the differences in the EATRs of European clusters, the following hypothesis is formulated:

H2: The effective average tax rate has a non-linear negative effect on the attraction of greenfield FDI in European economic clusters, with a higher effect at low tax rates and a lower effect when the initial tax rates are higher.

3.3 Tradeoff Theory

As described in Section 2.4, a tradeoff between FDI determinants is assumed, following the different policy decisions regarding their relative effect and implementation speed (Casi & Resmini, 2014). Based on these arguments and in line with this research, a tradeoff effect will be distinguished with the FDI determinants on two sides.

First, the corporate tax rates are considered. Tax rates are one of the largest threats to territorial competition and have a very fast implementation speed as an FDI determinant (Burger et al., 2013). Second, other FDI determinants are considered following the OLI paradigm. The regional characteristics of the political situation, global interconnectedness, and knowledge-based specialization are determinants with a relatively slow implementation speed and effects that can vary over time. In most cases, because changing these characteristics takes more time, the effect increases over a certain period and is more stable relative to abrupt changes in the tax rates.

Based on these arguments, the underlying relationship and interactions of the two groups of FDI determinants will be investigated. To do so, the relative weight of the determinants will be considered in different situations of the corporate taxes and the remaining characteristics (Goodspeed et al., 2007; Wijeweera et al., 2007). Thus, the relative weight will be estimated by investigating the effects at different rates of corporate taxes.

In the tradeoff, the non-linear relationship of corporate taxes on the likelihood of FDI inflows in a European cluster will be combined with the remaining FDI determinants that are expected to have a statistically significant impact. Following the research of Lawless et al. (2018), it is expected that the effect of corporate taxes will be relatively small when the tax rates are high and that the effect will be large(r) when the tax rates are relatively low. For the remaining determinants, importance is expected to increase over time and be more stable than taxes. Thus, in a situation where taxes are less important, the relative performance of the remaining determinants is expected to increase.

Hence, the other FDI determinants are expected to be less important in a situation of low corporate taxes and more decisive in a situation of high(er) corporate taxes. Combining the underlying relationship of the FDI determinants and the situation of non-linearity, the following hypothesis is stated:

H3: The underlying relationship of the FDI determinants shows a tradeoff effect in the attraction of greenfield FDI towards European economic clusters, with a higher but diminishing effect, of corporate taxes and an opposite effect in the remaining FDI determinants, as the tax level increases.

4 Data

4.1 Greenfield FDI

The data used to conduct this research originates from multiple sources. The main dataset is retrieved from the Financial Times fDi Markets database and contains information on cross-border greenfield investments on a worldwide level over the period from 2003-2018 (Financial Times, 2022). This dataset will be filtered on Europe using the countries of the European Economic Area (EEA) and including Switzerland. The cluster classification of this research, metropolitan NUTS 3 regions, represent all agglomerations of at least 250.000 inhabitants (Eurostat, 2020, 2021). Each agglomeration includes one or more NUTS 3 regions based on the population per cluster, creating 255 European metropolitan clusters and 33,368 greenfield investments for 2003-2018. Furthermore, since data is available on the regions of 16 years, there are 4,080 total observations of investment possibilities over the regions.

Figure 1 presents an overview of the geographical distribution of the greenfield investments observed in the fDi Markets database. The figure is made by matching NUTS-3 regions to the metropolitan regions and their received investments using QGIS. The outlined areas represent the metropolitan clusters defined by the classification, located in the EEA or Switzerland. The remaining grey areas do not meet the restriction of being a cluster and are, therefore, not in the scope of this research. The yearly averages are displayed to show the investment spread across Europe. Noticeable are the regions in the northwest of Europe, which are relatively close to each other due to a higher population density. The average number of yearly investments for the 255 regions is 8.18, with a standard deviation of 17.22. Of the 255 regions, 39 clusters (15%) have no average yearly investments. Sixty-four regions (25%) have an average yearly investment of one, and 67 regions (26%) have 2-4 yearly investments. The groups with more investments, between 5-9, and the regions with 10+, have 37 (15%) and 48 (19%) average yearly greenfield investments, respectively. The distribution is more skewed to the left for the total investment numbers over the entire period of 2003-2018. In this case, the largest groups are the regions that did not receive an investment in a specific year and the regions that received only one investment in a specific year, 1,146 (28%) and 695 (17%), respectively. The six regions with the most attracted FDI received over 1,000 greenfield investments over the full period – Paris, Dublin, Madrid, Amsterdam, Frankfurt am Main, and Barcelona. Together, they account for over 25% of the total greenfield investments.

4.2 Variables

The remaining data for the hypotheses is retrieved from various sources. These datasets contain information on the remaining corruption hypotheses, including tax rates and FDI determinants at the country- or regional level. Therefore, no observations will be lost due to the cluster classification. The main variables are stated below:

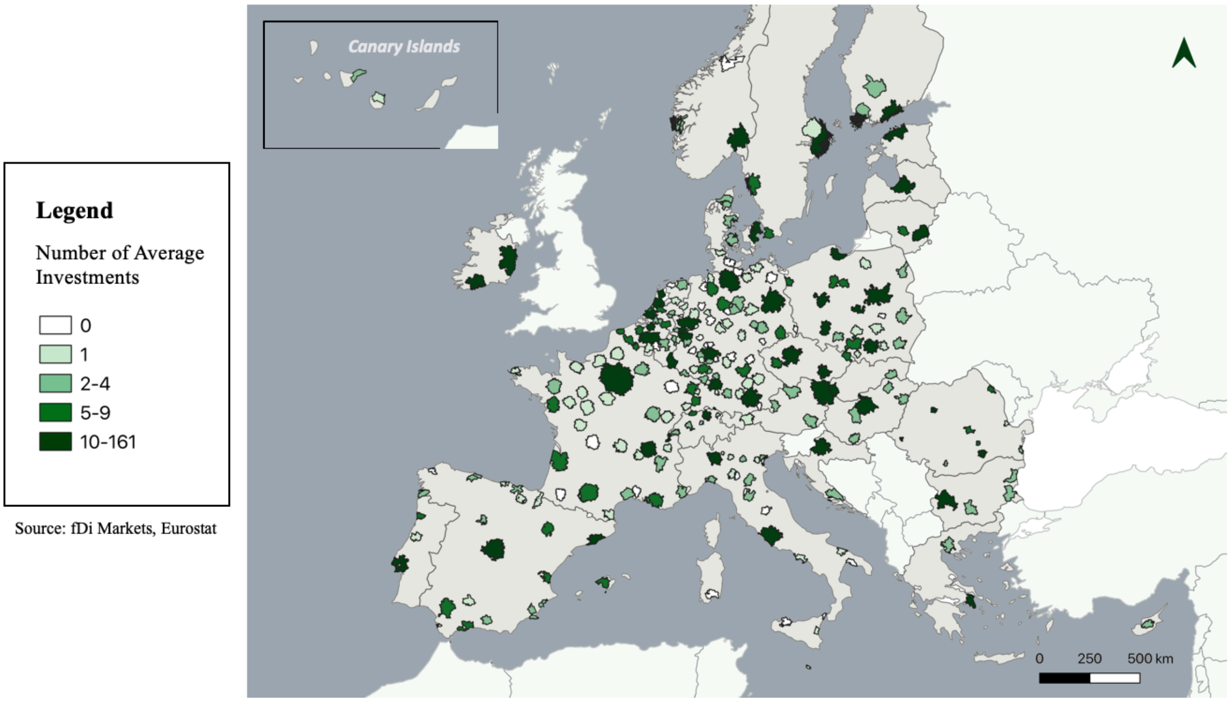

Figure 1: Number of yearly average greenfield investments per European metropolitan cluster (2003-2018)

Corruption Perceptions Index

The Corruption Perceptions Index (CPI) is part of the European Sustainable Development Goals and monitors progress on peace, justice, and strong institutions (Transparency International, 2012; Eurostat, 2022a). The indicator scores and ranks countries based on how corrupt a country's sector is perceived. On this scale, a score of "0" represents a high level of corruption, and a score of "100" represents a very clean country with a low level of corruption. This data is available on a country level and added to the research from 2003 to 2018. As described in Section 3.1, it is expected that corruption deters FDI. Since a higher CPI score represents a cleaner country with less corruption, the variable is expected to have a positive effect on the number of attracted greenfield investments.

Global Interconnectedness

Global interconnectedness is related to physical infrastructures that facilitate the mobility of people and goods. In order to capture this in a variable, this research considers multiple accessibility indicators. Spiekermann and Wegener (2006) describe how the potential accessibility relies on the assumption that the attraction of a location can increase with its GDP and decreases with different characteristics, such as distance, costs, or travel time. Furthermore, the accessibility indicators of multimodal transport are linked and positively correlated with each other.

Spiekermann and Wegener (2006) calculate the potential accessibility for four different methods of transportation; road, rail, air, and multimodal.[1] To calculate this number, the population in the destination region is weighted by the travel time to go there. The weighted population is summed up to the indicator value for the accessibility potential of the origin region (Spiekermann & Wegener, 2006). These factors are expressed as an index related to the ESPON average.[2] Furthermore, the variables are estimated in three different years throughout the investment period (2006, 2011 and 2014) due to low variation in the travel times, and are retrieved from the ESPON database (2023). This research expects that the attractiveness of a metropolitan cluster for greenfield FDI increases with its potential accessibility.

Specialization

The specialization variable uses the industry sector division of the fDi markets database. This database categorizes investments in multiple industry sectors. This variable defines knowledge-intensive sectors based on classifying knowledge industries using NAICS codes. Industries are knowledge-intensive based on their knowledge producers, R&D, and percentage of high knowledge worker inputs, such as technology, healthcare, and financial services (Statistics Canada, 2022). Table 5 of Appendix A presents an overview of the knowledge-intensive sectors.

Furthermore, a region overview visualizes the main industry activity as knowledge-based if the region had the most yearly investments in one of these sectors from 2003 to 2018. The yearly approach will create a binary variable with the value of "1" if the region received more investments in a knowledge-based sector than in the remaining (labor-intensive) sectors for each year. Based on the information presented in Section 3.1, it is expected that the knowledge specialization attracts greenfield FDI towards a cluster.

EATR

As described before, the effective average tax rates are the effective rates paid by the corporations. This rate differs from the statutory tax rates because of tax deduction possibilities, causing differences between corporations, regions, and countries. Hence, the EATR is generally lower than the statutory taxes and fluctuates more between regions (Lawless et al., 2014, 2018; Janský, 2019; Devereux & Griffith, 2002, 2003). The EATR variable is retrieved on a country level from the European Commission (2022) and is available from 2008-2018. Next to this numerical variable, tax dummies categorize the

[1] There is no available data on potential accessibility over water as it is not considered by Spiekermann & Wegener (2006).

[2] Since the calculation is the same for all transport methods, the three separate indicators (especially road and rail) may be highly correlated because of similar travel times. The correlation will be estimated, and (if needed) the models will be corrected in Section 5.5.

EATR level in a country. A high–medium–low scenario presents that the EATR high variable will have the value of '1' if the EATR is larger than 30%, the EATR medium will be '1' if the EATR is between 20 and 30%, and lastly, the EATR low scenario will be '1' if the EATR is below 20%.

Control variables

Furthermore, this research includes several variables to control for various remaining effects. The control variables will be collected from 2003 to 2018 and added on different levels, ranging from the metropolitan region to the country level.

First, data is added on a metropolitan level to collect additional information on the regions. The variables are obtained from Eurostat (2023b). The demand factors, measured by the local market size and the accessibility to the market, are partly captured in the main variables of this research (in the accessibility and sector specialization). Therefore, only the market size, measured as the local GDP of the region, will be added as a demand factor. This variable is expected to positively affect the number of greenfield investments in the region. Furthermore, factors in external economies include positive externalities from locating in a metropolitan region (McCann & Shefer, 2004). The population density, measured as the population per square kilometer, can indicate more urbanized areas and can therefore affect the investment decisions of MNEs. Hence, this variable is expected to positively affect the number of greenfield investments. However, the workforce level, measured as the unemployment rate, can indicate the absence of agglomeration effects in metropolitan regions and is therefore expected to have a negative effect on the number of investments.

Second, the paper adds data on the country level to collect information on the supply factors. Following Burger et al. (2021), these factors consist of production costs and quality. For efficiency-seeking motives, the labor costs are added and measured as the nominal compensation per employee. This variable is retrieved from the AMECO database of the European Commission (2023) and is available annually. The labor costs are expected to be linked to the profitability of FDI. Since higher labor costs decrease the profits regarding FDI, this variable is expected to negatively influence the number of attracted investments in a region (Spies, 2010). The cost of capital is captured by the taxes, measured as the statutory corporate tax rate. In line with the aim of this research, the statutory taxes will be used to control or visualize the differences with the EATR. Lastly, the country's annual inflation rate is added to capture consumer price developments (The World Bank, 2023). This variable is expected to negatively influence the investments of MNEs regarding their location decisions.

Below, an overview of the main and control variables is added in Table 1. The table visualizes the determinants with their variable names, corresponding proxy, and their expected effect (sign) on the number of attracted greenfield investments in the metropolitan regions. Table 6 of Appendix B reports the corresponding descriptive statistics.

Table 1: The Determinants of Greenfield FDI for European Clusters

Variable	Proxy	Expected Sign
Main Variables		
Corruption Perceptions Index	Score of the country's corruptness (0-100)	+
Global interconnectedness	Potential accessibility in rail, road, air of multimodal transportation (ln)	+
Knowledge sector	Specialization of the main activity in the knowledge sector	+
Effective Average Tax Rate (EATR)	Total effective average tax rates as paid by the corporations (%)	-
Control Variables		
Market size	Local GDP (ln)	+
Population density	Population per square kilometer (ln)	+
Workforce	Unemployment rates (%)	-
Policy rates	Statutory taxes (%)	-
Labor costs	Nominal compensation per employee (ln)	-
Inflation	Consumer price index (%)	-

NOTES: This table shows the determinants of greenfield FDI which are the factors of interest for the empirical analysis. The variables are shown, together with their proxies and expected signs.
Sources: Eurostat (2022a, 2023b, 2023c), European Commission (2022, 2023), The World Bank (2023)

5 Methodology

Next, the research introduces the econometric approach used in the analysis. As discussed before, the available data consist of greenfield investments in European clusters and regional characteristics for the period of 2003-2018. The dependent variable is the investments in greenfield FDI. As described in Section 4.1, this variable represents the number of investments for each metropolitan cluster for every investigated year. Therefore, it takes on relatively few values, including zero, and cannot have a normal distribution (since this is for continuous variables that can take on all values). If the variable takes on very few values, the distribution can be very different from normal (Wooldridge, 2015). Hence, linear probability models are outside scope and count data models are of interest to analyze the relationship between the number of investments and the regional characteristics. The different tests used for the hypotheses will be discussed below.

5.1 Count Data Models

Certain nonnegative dependent variables can be categorized as count variables and can take on integer values such as {0, 1, 2,…}. Count variables cannot have a normal distribution and take on relatively few values. Instead, the nominal distribution for count data is the Poisson distribution (Wooldridge, 2015).

The formula for this probability mass function with y as random variable that indicates the number of times an event occurred, is given by:

$$(1)\ Pr(y \mid \mu) = \frac{e^{-\mu}\mu^{y}}{y!} \text{ for } y = 0, 1, 2, \dots$$

and indicates the probability of a count. $\mu > 0$ is the sole parameter defining this distribution, the mean of the function, and an increase of this parameter shifts the distribution to the right. μ is also the variance. If this is equal to the variance (Var(y) = μ), there is equidispersion, and if the mean is larger than the variance, there is overdispersion. The probability mass function is linked to the most basic model with count variables, the Poisson Regression Model (PRM). This model is a special case of the Negative Binomial Regression Model (NBRM). Furthermore, the Poisson process assumes that the events are independent, meaning that the occurrence of an event does not affect the probability of the occurrence of an event in the future. However, the variance in the model is often larger than a Poisson process would suggest. This can be caused by observed heterogeneity (differences among sample members), where μ differs across individuals and results in a pattern of over-prediction of counts which are ≥ 1 and under-prediction of 0s.

Although the PRM accounts for observed heterogeneity, it rarely fits in practice because of overdispersion. This causes the estimates to be inefficient and results in biased standard errors. The NBRM adds a parameter ($\tilde{\mu}$) to allow for the conditional variance of y to exceed the conditional mean. In this model, there is also variation due to unobserved heterogeneity. Therefore, the conditional mean is still μ, but the variance is larger because of the error term.

5.2 Panel Data for PRM and NBRM

Testing for and choosing between the models can be done by testing for overdispersion. Using a one-tailed test for α, the dispersion parameter, in a negative binomial regression model. A significant statistic indicates overdispersion and highlights NBRM as the preferred model. In the greenfield FDI dataset, the tests show a variance in the models that is substantially larger than the mean.[3] Hence, the use of the negative binomial model seems justified (Long & Freese, 2014; Williams, 2021).[4]

However, the count of investments is annually per region from 2003-2018. This suggests using panel data and brings other restrictions and model-fit tests. Using yearly averages in cross-sectional data is also considered, such that each metropolitan cluster has one observation. However, the data is already

[3] The conditional variance equals the squared term of the standard deviation. For the number of greenfield investments, the conditional variance is 48 times larger than the conditional mean (Table 6 of Appendix B).

[4] This is supported by the model fit prediction where the Poisson predicts less zeroes than observed zeroes and the NBRM method fits the data better. See Figure 4 of appendix D.

filtered on metropolitan clusters, which have a relatively high probability of receiving at least one investment due to their positive effects in agglomeration economies (Franca et al., 2021). This results in observed heterogeneity with a skewed distribution and under-observation of '0' counts. Thus, in line with the aim of this research of finding underlying effects of the FDI determinants in agglomeration economies, fixed effects in panel data are considered.[5]

Transforming the data of the NBRM in a panel data estimation is described as fitting a conditional fixed effects model. However, this is not a true fixed effects method as the likelihood estimator does not necessarily remove the fixed effects, where the fixed effects apply to the distribution of the dispersion parameter and not to the coefficient of the model. An alternative is the unconditional fixed effects model, which adds panel dummies to a conventional NBRM to estimate the fixed effects (Allison & Waterman, 2002; Guimarães, 2008; StataCorp, 2021). Regarding this research, yearly panel dummies can be added to control for economy-wide time-invariant effects. However, the unconditional fixed effects model for the NBRM poses two problems: a potential incidental parameters problem (IPP), producing inconsistent estimates of the coefficients, and a (potential) downward bias in the standard error estimates.[6] As a response, Allison and Waterman (2002) find no evidence of the incidental parameters problem in their study on the unconditional model and the (potential) downward bias in the standard error estimates can be corrected using goodness-of-fit statistics.

Another possibility is using a fixed effects Poisson model. This model does not suffer from the incidental parameters problem and is fully robust in the sense that it only needs the structural conditional mean assumption for consistency in the distribution. However, using the Poisson model transformation is restricted since the data in this research suffers from overdispersion. Allison and Waterman (2002) investigate both options (the unconditional negative binomial estimator and the fixed-effects Poisson estimator) and show that the coefficients are somewhat similar.

In conclusion, following Allison and Waterman (2002), this research uses the unconditional negative binomial regression model by adding year panel dummies to control for economy-wide time invariant effects. In addition, the model is challenged in a different setting by adding a fixed effect Poisson estimator to see if the results are similar (see Table 9 of Appendix X).

[5] Random effects are not considered as modeling intracluster correlation (correlation of observations in the same cluster) is not in scope and it is not in line with the substitute Poisson model for the NBRM with panel dummies, which works with fixed effects (StataCorp, 2017).

[6] The IPP arises when a model with many group-specific parameters (such as individual fixed effects) is estimated. This often occurs in non-linear panel data with fixed effects because the fixed effects can be estimated in a linear setting (Lancaster, 2000).

5.3 Negative Binomial Regression Model

In the NBRM model, the distribution of observations is still Poisson, including the parameter that allows the conditional variance of y to exceed the conditional mean. This $\tilde{\mu}$ adds error term ε to the independent variables which is assumed to be uncorrelated with the x's. δ defines as $\delta = e^{\varepsilon}$. This results in the following function for $\tilde{\mu}$:

$$(2)\ \ Pr(y_i \mid x_i, \delta_i) = \frac{e^{-\tilde{\mu}_i}\tilde{\mu}^{y_i}}{y_i!} \quad \text{for } y_i = 0, 1, 2, \ldots, i \in \{1, \ldots, N\}$$

In line with the aim of this research, let Y_i denote the count of greenfield investments in region i over the period of 2003-2018, with $i \in \{1, \ldots, N\}$ and N the total number of European metropolitan clusters. Y_i is a nonnegative integer and follows a negative binomial distribution with conditional mean μ_i and overdispersion parameter α. Γ is the gamma function and x_i is the $p \times 1$ vector of explanatory variables for region i. Given μ_i and α, the probability mass function of Y_i is (NCSS, 2023):

$$(3)\ \ Pr(Y_i = y_i \mid \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i}, \ \ y_i = 0, 1, 2, \ldots, i \in \{1, \ldots, N\}$$

The parameter μ_i may be interpreted as the risk of a new occurrence of the event during a specified exposure period. Furthermore, the conditional mean is considered to be equal to $e^{(c + \beta' x_i)}$, where c is a constant and β the $p \times 1$ vector of regression coefficients.

The interpretation of the results of the NBRM can be done using the marginal effects and the Incidence rate ratios (IRR). These methods differ slightly. The marginal effects generally measure the impact that a change in one variable has on the dependent variable with all other variables held constant. The IRR coefficient explains how the changes in the independent variable affect the rate at which the dependent variable occurs. In line with this research and its hypotheses, both will be estimated and interpreted.

5.4 Non-linearity of Taxes

In the second hypothesis, the EATR and its non-linear effect are central. To test the non-linear characteristics, multiple methods will be applied and compared. The first method categorizes the regions into different groups based on their EATR (high – medium – low). This creates a non-linear model with multiple coefficients for the tax rates. However, the variance and standard error are expected to increase, impacting the validity of the estimation and interpretation of the data in a negative way. The second method is in line with Lawless et al. (2018). It includes, next to the EATR variable, the squared term of the EATR to account for potential non-linearity in its relationship with the number of investments. This method expects that the tax rate has a negative effect on the number of investments and that the squared term has a small positive effect, such that the tax rates have a decreasing effect on the investments at higher rates of the EATR. The third method of testing the non-linearity of taxes estimates direct effects in the functional form. Adding and estimating polynomials predicts the fit of the model. By plotting the model's residuals against the tax rates, the model's fit will be found if the residuals are randomly distributed. Lastly, the most accurate model will be estimated and interpreted to test the hypothesis for the non-linearity of the taxes.

5.5 Tradeoff Methods

The last hypothesis of this research investigates the tradeoff between taxes and FDI determinants. After stating the non-linearity of taxes and its best-fitted model, this will be implemented in the preferred model of Hypothesis 1. This creates the unrestricted model combining the variables from the first two hypotheses. To estimate the underlying relationships, the unrestricted model adds interaction terms between the EATR and the remaining FDI determinants. The changes relative to the EATR will be interpreted using the estimations of marginal effects at different tax rate levels. The interpretation will be made for all main variables included in the unrestricted model.

As stated before, it is expected that the effect of taxes will be low when tax rates are high and that the effect will be larger when the tax rates are lower. Estimating and plotting the margins of the main variables against the tax rates is expected to find a threshold point where the effect of increasing taxes becomes less important and where the effect of the remaining FDI determinants grows. To interpret the effects of the best fitting model, the marginal effects will be estimated at different means of the EATR. Two methods will be used to estimate this tradeoff relationship. (1) the marginal effects of a high–medium-low scenario will be estimated to interpret the coefficient and significance. (2) the marginal effects will be plotted at different EATR levels to conclude the main determinants' curves.

5.6 Multicollinearity and Robust Standard Errors

Regarding collinearity, this research includes correlation tables and standard errors to investigate and detect potential collinearity issues.[7] Signs such as increasing standard errors, wide confidence intervals, and highly correlated variables in the models can be caused by multicollinearity (Williams, 2015). In general, (multi)collinearity is assumed to pose no statistical problems when the interpretation is done within the context of the research. Since the characteristics that attract inward FDI to metropolitan clusters are intuitively correlated (such as the accessibility variables), a slight increase in the correlation factors is expected (Belsley et al., 1981; Mansfield & Helms, 1982). However, the potential accessibility

[7] Variance inflation factors (VIFs) are also considered to detect potential collinearity issues (Robinson & Schumacker, 2009). However, VIFs will not correctly identify the effect of the potential multicollinearity on the standard errors for non-linear models like the NBRM since the multicollinearity among predictors does not inflate the coefficient of the standard errors in the same way as it does in linear models.

by road and rail is too highly correlated (over 0.96). Therefore, these variables will not be included in the same model. The same principle holds for the EATR and statutory taxes as they measure similar parameters. Appendix C presents information on (multi)collinearity, with the correlation overview in Tables 7-8.

Furthermore, the paper implements robust standard errors using the empirical variability of the model residuals. This equals the difference between the observed and predicted outcome, computed based on the Huber-White estimator, and therefore explains the independence of the outcome (University of Virginia, 2020). Furthermore, robustness checks will be added to the methods to test if the models produce the same outcome in a different setting under different circumstances. The robustness will be added in Section 7 to explore the sensitivity of the results.

5.7 Model Selection

This section presents the model selection for the hypotheses. Since the model selection differs for each hypothesis, the models will be discussed separately. Furthermore, since competing model specifications within the negative binomial regression model exist, the best model will be selected based on favorable goodness of fit parameters. Therefore, the Akaike (AIC), Bayesian (BIC), and (adjusted) McFadden's R^2 parameters will be estimated. Appendix E presents the model choice estimators and arguments.

Based on the presented arguments for the best model fit, the first hypothesis includes the CPI score for part (a), the potential accessibility variables for road and air for part (b), and the knowledge sector specialization dummy for part (c). This model (specification 3 of Table 10) is selected as the best-fitted model based on the highest adjusted McFadden's R^2 and the lowest AIC and BIC values.

The second hypothesis tests the model specifications based on their goodness-of-fit parameters and highlights the non-linearity of taxes by plotting the randomization of the residuals of the functional form. Since the goodness-of-fit estimators are relatively close to each other, the choice criteria are on the standard error and distribution on the residual plot. Based on these estimation methods, the EATR model, including its squared term, is selected as the best-fitted model (model specification 3 of Table 12). Figure 2 presents the distribution of the residual plot. Although most residuals are centered around '0' because of outliers, there is no clear trend of residuals, and the distribution is random. This is similar to the residual plots of the remaining model specifications of Table 12, meaning that none of the model specifications suffer from linear residuals (Figure 5 of Appendix E).

The third hypothesis is an unrestricted model combination of Hypothesis 1 and 2. This last model combines the best-fitted model specifications. Therefore, the negative binomial regression model will consist of the main variables CPI score, the three separate potential accessibility parameters, the specialization in a knowledge sector, and the EATR with its squared term. The functionality of this model will be compared to the two restricted models of the previous hypotheses.

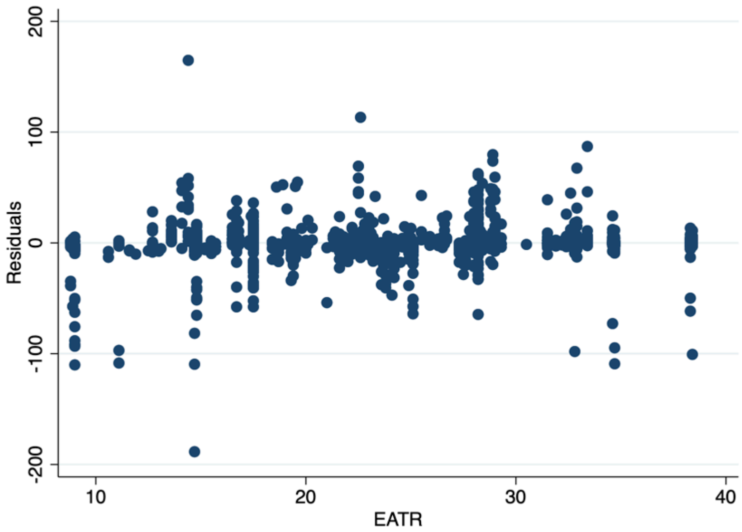

Figure 2: Residual plot of the best-fitted EATR model

Notes: This figure plots the residuals for the best-fitted model of Table 12 Appendix E (model specification 3). A random distribution of the residuals is shown of the model with the EATR and its squared term.

6 Results

Next, the results of the empirical analysis are presented, using the data of metropolitan clusters on a regional and national level. Section 6.1 presents the analysis of the determinants of greenfield FDI. Section 6.2 examines the non-linearity of taxes and their relations with greenfield investments. Section 6.3 explores the tradeoff model using interaction terms and the marginal effects at different means.

6.1 Determinants of Greenfield FDI

Table 2 presents the estimation results of the negative binomial regression models for the number of greenfield investments in the metropolitan clusters. This section examines the best model fit of the first hypothesis to interpret multiple results using the marginal effects and Incidence rate ratios. The coefficients of the negative binomial regression of this model are presented in model specification (3) of Table 10 Appendix E. The coefficients of the control variables in Table 2 are, except for the unemployment rate, all significantly related to the number of attracted greenfield investments in a region. The local GDP, population per square kilometer, and compensation per employee are in line with the expectation from Table 1. Not in line with the expectations is the inflation rate. A possible explanation could be that because inflation is measured at the country level, certain clusters could be outliers based on their large amount of attracted investments while having a relatively high inflation rate at the country level.

Table 2: Negative Binomial Model for FDI determinants on the Number of Greenfield Investments

	Average Marginal effects	Incidence rate ratios
Main variables		
CPI Score	0.102***	1.011***
	(0.023)	(0.002)
Potential accessibility by road (ln)	-1.720***	0.837***
	(0.386)	(0.033)
Potential accessibility by air (ln)	1.432**	1.160**
	(0.760)	(0.091)
Knowledge sector specialization	4.217***	1.546***
	(0.433)	(0.065)
Control variables		
Local GDP (ln)	12.082***	3.487***
	(0.582)	(0.109)
Population per km^2 (ln)	1.037***	1.113***
	(0.328)	(0.037)
Unemployment (%)	-0.054	0.994
	(0.046)	(0.005)
Compensation per employee (ln)	-16.039***	0.191***
	(0.886)	(0.012)
Inflation (%)	1.130***	1.124***
	(0.176)	(0.020)
Constant		0.005***
		(0.002)

Notes: This table presents estimation results of the negative binomial model for the best-fitted model regarding FDI determinants. The first model presents the estimates of the marginal effects and the second model presents the Incidence rate ratios. The coefficients and the yearly panel dummies to capture the time invariant effects are added in Table 10 and 11 of Appendix E, respectively. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

The yearly dummies, as presented in Table 11 of Appendix E, show positive and significant results for the years between 2013 and 2016.[8] This indicates a slight increase in the number of investments relative to the start of the period, 2003, which is in line with the slightly increased number of investments over the years.

Regarding the main variables in this model specification, it is noticeable how the CPI score, knowledge specialization, and potential accessibility by air are positively and significantly related to the number of greenfield investments. The accessibility by road is negatively related to the amount of attracted FDI.

Based on the results of Table 2, multiple conclusions can be drawn for the sub-hypotheses of Hypothesis 1. The marginal effects and Incidence rate ratios are estimated for interpretation and

[8] Only the coefficients of the panel dummies are added in Table 11 to present the sign and significance of the variables. The panel dummies' marginal effects and incidence rate ratios are not presented as they will not be interpreted.

comparison purposes. The interpretation of the marginal effects tells us that, after controlling for other variables, on average the knowledge sector specialization attracts 4.217 more greenfield investments than regions that are not specialized in the knowledge sector. For the CPI score, this value is, on average, 0.102 more. For potential accessibility by road and air, this is respectively 1.720 less and 1.432 more investments, after controlling for other variables. The negative value for the potential accessibility of road is not in line with the expectations, and it can be argued that since Europe is so well connected over the road, other determinants of accessibility are more decisive (World Economic Forum, 2013). Thus, increasing the potential accessibility by road will not lead to investments as it does not improve the accessibility to, for example, large business areas (under the assumption that this infrastructure is already present in Europe).

The incidence rate ratios explain more intuitively how the changes in the independent variables affect the rate at which the greenfield investments occur (or, in other words, present the estimated rate ratio for a one-unit increase in the number of greenfield investments). Therefore, the IRR is interpreted for each sub-hypothesis. Starting with part (a) of this hypothesis, stating that corruption deters the attraction of greenfield FDI for the metropolitan clusters, the CPI score visualizes the country's corruptness with a score of 100, presenting a 'clean' country. Therefore, this is expected to have a positive effect on the number of investments. For the CPI score, it can be expected that a metropolitan cluster with a CPI score of one point higher than another cluster receives 1.1% more greenfield investments, ceteris paribus. This result provides support for the first sub-hypothesis, stating that corruption deters inward FDI. Similar to the CPI score, a European metropolitan cluster with a (natural log of the) score for potential accessibility by air that is one point higher than another cluster is expected to receive 16.0% more investments, ceteris paribus. However, for the potential accessibility by road, a metropolitan cluster with a (natural log of the) score for potential accessibility by road that is one point higher than that of another cluster is expected to receive 16.3% less investments, ceteris paribus. As a result, the findings for the accessibility variables are contradictory. Thus, there is not enough support for the second part (b) of the first hypothesis, which states that a high degree of global interconnectedness attracts inward greenfield investments. Lastly, the knowledge-based specialization of the clusters' main activity is interpreted. From Table x, it can be concluded that a European metropolitan cluster with a knowledge sector specialization compared to a cluster that does not have this as its main activity is expected to receive 54.6% more greenfield investments, ceteris paribus. This provides evidence for the last part of the first hypothesis, (c), which states that greenfield investments are attracted to knowledge-based sectors.

In conclusion, the first hypothesis presents evidence that supports sub-hypothesis (a) and (c) because of a positive effect of the CPI score and knowledge sector specialization on the number of attracted greenfield investments by metropolitan clusters. However, since the results of the potential accessibility of road and air are contradictory, there is not enough evidence to support the positive effect of the potential accessibility on FDI (sub-hypothesis (b)).

6.2 Greenfield FDI and Effective Average Tax Rates

The second part of this research focuses on the relationship between corporate taxes and the attraction of greenfield FDI towards European metropolitan clusters. As described in Section 5.3, the non-linearity of taxes is tested using multiple econometric methods. The best model fit (specification (3) of Table 12 in Appendix E) consists of the EATR and its second-order polynomial. Table 3 presents the marginal effects and IRR estimators of this model.

Table 3: Negative Binomial Model for the EATR on the Number of Greenfield Investments

	Average Marginal effects[9]	Incidence rate ratios
Main variables		
EATR	-0.715***	0.813***
	(0.076)	(0.021)
$EATR^2$		1.003***
		(0.000)
Control variables		
Local GDP (ln)	14.880***	3.930***
	(0.730)	(0.115)
Population per km^2 (ln)	1.765***	1.176***
	(0.371)	(0.038)
Unemployment (%)	0.052	1.005
	(0.047)	(0.004)
Compensation per employee (ln)	-10.261***	0.389***
	(0.799)	(0.024)
Inflation (%)	0.639***	1.061***
	(0.215)	(0.021)
Constant		0.003***
		(0.001)

Notes: This table presents estimation results of the negative binomial model for the best-fitted model regarding the EATR. The first model presents the estimates of the marginal effects and the second model presents the Incidence rate ratios. The coefficients and the yearly panel dummies to capture the time invariant effects are added in Table 12 and 13 of Appendix E, respectively. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

The control variables in the presented model have similar signs and effects as the variables in Table 2. The interpretation is as follows for the marginal effects and IRR of this hypothesis. Regarding the marginal effects and after controlling for other variables, on average increasing the EATR with one unit generates 0.715 less greenfield investments than regions that do not follow this increase. For the IRR, a metropolitan cluster with an EATR one point higher than another cluster is expected to receive

[9] The marginal effects of interaction terms are not calculated because it cannot change independently of the values of the component terms. Thus, a separate effect for the squared (interaction) term cannot be estimated. The IRR does not suffer from this and provides interpretation possibilities for the squared term (Buis, 2010; Williams, 2012).

18.7% less greenfield investments, ceteris paribus. For the squared term of the EATR, a positive effect is found, such that a metropolitan region with one point higher than another region is expected to receive 0.3% more investments, ceteris paribus.

The small positive effect of the squared term indicates that the taxes have a decreasing effect on the investments at higher rates of the EATR. This result is in line with the research of Lawless et al. (2018), who found a negative and significant effect for corporate taxes and a positive and significant squared term, indicating that the strength of the negative effect of taxes moderates as the rate of tax increases.

This provides evidence for the second hypothesis, which discusses the non-linearity of the effective average tax rate on the attraction of greenfield FDI in European economic clusters, with a higher effect at low tax rates and a lower effect when the initial tax rates are higher.

6.3 Tradeoff Effects and Greenfield FDI

The results discussed so far discussed separate restrictive models, including main variables regarding FDI determinants and the EATR (including the EATR and its second-order polynomial). As motivated in section 2.4, there is reason to suspect that the attraction of greenfield FDI towards European metropolitan clusters differs across the determinants. Therefore, a tradeoff effect is examined on the underlying relationships between these determinants. The tradeoff investigates the relationship of FDI determinants in agglomeration economies with the corporate taxes on the one side and the remaining FDI determinants on the other. It assumes that the weight of the taxes and other FDI determinants influences the attractiveness of certain clusters in multiple ways, given a different set of regional characteristics. Combining the results from the first two hypotheses, an unrestricted model will be examined, including the main variables of the best-fitted models from Hypotheses 1 and 2.

Furthermore, interaction terms will be added to the unrestricted model to correctly estimate the effects of the FDI determinants at different levels of the tax rate. Although the marginal effects in nonlinear models are not constant over the entire range (even in the absence of interaction terms), they cannot fully estimate the effects between the independent variables (Karaca-Mandic et al., 2012). Therefore, this section adds interaction terms to expand understanding of the underlying relationships among the main variables. In this case, it is expected that the EATR and the remaining FDI determinants influence each other. With the goal of the tradeoff to identify the effect of high and low tax rates on the remaining FDI determinants, interaction terms are added to the unrestricted model for each of these determinants with the EATR.

Regarding the fit of the unrestricted models, Table 14 in Appendix F presents estimates that highlight the unrestricted models as the best model fit compared to the first two hypotheses. With a higher adjusted McFadden's R^2 and lower AIC and BIC statistics, the models present the best goodness-of-fit parameters. Furthermore, since the EATR with its second-order polynomial is added, the choice criteria of the nonlinearity in the taxes (random distribution of residuals and the fluctuations in standard errors) have already been considered. Therefore, unrestricted models are preferred over restricted models. When comparing the unrestricted models (the model without interaction terms and the model including the interaction terms), the goodness-of-fit parameters are not decisive. Examining the significance of the coefficients is the solution to check if the model with interaction terms is preferred. Regarding this investigation, model specification (3) of Table 14 in Appendix F has significant interaction terms for air and knowledge with the EATR and is therefore preferred.

After stating the preferred unrestricted model with interaction terms, the tradeoff model examines the underlying relationships of different tax levels. The expectations of the EATR itself are in line with the finding of Lawless et al. (2018) that the effect of the EATR is relatively small when the taxes are high and that the effect will be higher when taxes are relatively low. For the remaining FDI determinants, the tradeoff expects that the effects are higher at high levels of taxes and lower at low tax levels. Table 4 investigates the underlying relationships by estimating the effects at different means of the EATR. This is divided into three different groups, following the means of the EATR high-medium-low scenario. The tested levels of the EATR are 35% for high, 25% for medium, and 15% for the low scenario.

Table 4: Marginal effects of Main Variables in a High – Medium – Low Scenario for the EATR

	High	Medium	Low
Main variables			
CPI Score	0.142^{***}	0.180^{***}	0.409^{***}
	(0.047)	(0.021)	(0.094)
Potential accessibility by road (ln)	0.482	-0.103	-1.367
	(0.590)	(0.351)	(0.134)
Potential accessibility by air (ln)	3.385^{***}	1.896^{**}	0.521
	(1.214)	(0.792)	(1.960)
Knowledge sector specialization	3.379^{***}	2.664^{***}	3.447^{**}
	(0.615)	(0.382)	(1.383)
EATR	0.231^{**}	-0.311^{***}	-1.669^{***}
	(0.108)	(0.051)	(0.395)

Notes: This table presents estimation results of the marginal effects of the main variables at different levels of the EATR. The marginal effects are based on model specification (3) from Table 14-15 of Appendix F The high scenario is estimated at an EATR level of 35%, the medium scenario at 25% and the low scenario at 15%. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Following Table X, the variables can be discussed separately. Starting with one side of the tradeoff, the EATR, it is visible how the effect is negative and significantly larger at low levels of the EATR and that this effect decreases as the tax rate increases to 25%. Eventually, the marginal effects are positive at an EATR level of 35%, but the coefficient still decreases in variation from 0. Therefore, diminishing marginal effects are noticeable at increasing levels of the EATR.

The CPI score also predicts significant effects in each scenario. This effect is relatively large at the low scenario for EATR and decreases significantly as the tax level increases. The knowledge sector specialization shows a potential U-shaped effect in the marginal effects, with larger effects at 15% and 35% and a smaller effect at 25%. The potential accessibility by road shows a similar relationship. However, these coefficients are insignificant. The potential accessibility by air shows an increasing effect as the tax rate increases from 25% to 35%. Although accessibility increases from 15% to 25%, this effect is insignificant. Table 16 of Appendix G presents similar effects, where a high–medium–low scenario is tested at different rates of the EATR.

Regarding the standard errors of the determinants, the lowest values are presented at the medium scenario as this is closer to 0 regarding the effects on the greenfield investments and estimated with higher validity.

To sum up, there is evidence presented in the marginal effects that suggest that the effects for the EATR are decreasing as the level of taxes increases, and that suggests that the effects of other variables such as the CPI score, potential accessibility by air and knowledge sector specialization increase significantly if the EATR increases to 35%. However, since the potential accessibility by air is only significant from 25% onwards and the knowledge specialization shows a possible U-shaped relationship, more evidence should be considered before concluding on the underlying relationship of the variables. Therefore, Figure 3 estimates the marginal effects for every EATR percentage point.

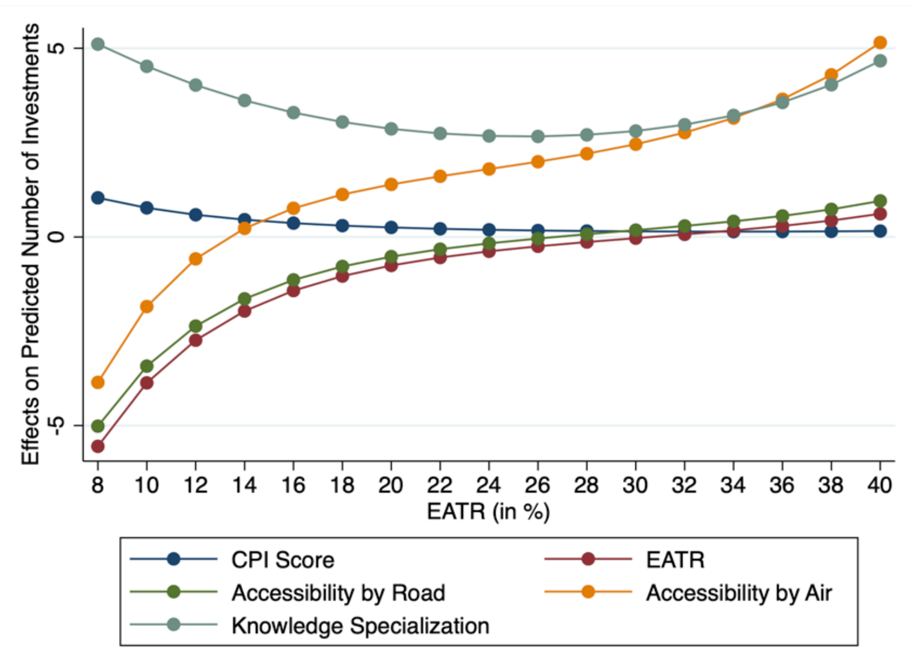

Figure 3: Marginal effects of the Unrestricted model on the Number of Investments at different EATR levels

Notes: This figure plots the marginal effects for the five main variables of Table x with the EATR on the x-axis and the number of greenfield investments on the y-axis.

Based on Figure 3, multiple conclusions are drawn. First, the EATR shows diminishing marginal effects with larger negative effects at lower levels of the EATR and eventually small positive effects at high levels of the EATR. Regarding the U-shaped relationship, this is visible in the knowledge sector specialization. The effects of the CPI score are smaller and, therefore, less visible. The potential accessibility by road seems similar to the EATR plot, but its effects are still insignificant. The potential accessibility by air shows another curve, visualizing that the marginal effects are larger at relatively low and high tax rates and smaller at medium tax rates.

Regarding the significance of the taxes, Table 16 and Figure 6 of Appendix G investigate this relationship further. Since no other significant effects are found for the potential accessibility by air at lower tax rates, the focus is shifted to the medium and high scenarios. This is also found in the confidence intervals of Figure 7, which are at their highest at low levels of the tax rate.

To sum up, the marginal effects of the EATR are significantly decreasing as the tax levels increase. For the potential accessibility by air, this effect is significant in the medium and high scenarios, showing increasing marginal effects as the tax levels increase. This points toward a tradeoff effect between the EATR and potential accessibility by air. For the remaining FDI determinants of the CPI score and the knowledge sector specialization, there is no tradeoff effect found. The CPI score decreases in line with the EATR itself over rising levels of taxes, and the knowledge specialization shows a U-shaped relationship. For the CPI score, this decrease in marginal effects could be because of the policy-related nature its shares with the EATR. At these higher levels of the EATR, policy decisions are affecting the business environment and other options could be more interesting (such as accessibility) due to the larger relative effects over time. For the knowledge specialization, this effect is U-shaped. Although this means that at higher levels of the EATR, its effect is increasing while the effects of the EATR decrease, it does not indicate a tradeoff since the opposite effect is not found at the lower values of the EATR. Lastly, the potential accessibility by road does not have significant marginal effects at all values of the EATR. This suggests that its effect is independent of the level of taxes because it is not significant regarding its relationship with the number of greenfield investments.

In addition, the next step in interpreting the unrestricted model with interaction terms is validating it by comparing the model to the restricted and unrestricted models without interaction terms. In comparison, it is noticeable that the unrestricted model with interaction terms yields the best model fit and captures the underlying relationship between the independent variables regarding greenfield investments towards metropolitan clusters. Figure 7 of Appendix G presents the results of the other restricted and unrestricted models.

To sum up, a tradeoff effect is found while investigating the underlying relationship between FDI determinants. For the potential accessibility by air and the EATR, it is stated that the marginal effects of the EATR are larger and significant at low levels of the EATR. When the EATR levels are increasing, the marginal effects of the EATR decrease, and the effects of the potential accessibility by air are opposite, such that it is significant and increasing from a tax rate of 15% onwards. This provides evidence for the third hypothesis, which states that the underlying relationships of FDI determinants show a tradeoff effect in the attraction of greenfield FDI towards European economic clusters.

7 Robustness Checks

This section offers robustness checks of the main estimates as a measure of the capacity of the results to remain unaffected by variation in the parameters of the research method. Therefore, robust methods are not influenced quantitatively by the examined factors and can indicate the generalizability and reliability of the results (Heyden et al., 2001). This section will discuss two methods of robustness checks. First, the use of alternative variables within the negative binomial regression is considered, with an important change being the tax rates of the statutory taxes. Multiple models will be made and challenged to the main model of this research. Second, the research will focus on the Poisson model with fixed effects to test if the results are similar in different settings following Allison and Waterman (2002).

7.1 Identification of Different Variables

In this section, the main model will be challenged using different observations and variables within the negative binomial regression. First, the statutory tax rates will be considered instead of the EATR. As stated in Section 4.2, these rates represent the cost of capital and are measured from 2003 to 2018. Therefore, more observations will be added to the model (since EATR data is available from 2008 onwards). The expectations of this model are similar for models including the EATR. To test this, the restricted model of the statutory taxes, including the remaining (main) variables, will be tested and compared to the unrestricted models, including the second-order polynomial and the interaction terms.

Furthermore, this section highlights the classification of the knowledge section specialization dummy. The classification for this variable identifies a specialization in the knowledge sector if a region has more knowledge-based investments than labor-intensive investments for each year from 2003 to 2008. However, this also includes regions with only one investment more in a knowledge-based sector than in a labor-based sector. This is closer to a predominance of investments in the knowledge-based sectors than it is to a specialization. A possibility would be to use another specialization description that only includes regions where the difference in knowledge- and labor-based investments reached a certain threshold or to dive further into the business sectors and activities, as many clusters have more than one specialization (Crozet et al., 2004; Burger et al., 2013). Therefore, a new knowledge specialization variable is created that recognizes a knowledge specialization as a region that received at least four investments, of which 75% are in a knowledge-based sector.[10]

Lastly, one of the control variables is characterized as the logarithm of the local region GDP in millions. Although this is on the logarithmic scale, the effects and influence of this variable are still large. To check whether the model is similar if this variable is less dominant, the change in GDP (in %) is added to the model. The discussed robustness variables are added to the descriptive statistics (Table 6 of Appendix B).

The different variables are systematically added to the models in Table 17 of Appendix H. For model specification (2), including the knowledge specialization threshold, interaction terms will be added to estimate the potential tradeoff effect.[11] Following this model choice, model specification (4) includes the interaction terms to estimate the potential tradeoff effect. This estimates significant tradeoffs with the statutory tax rates for the CPI score (in line with Section 6.3) and the potential accessibility by road and air. The potential accessibility by road differs from earlier results, which is expected to result from the added observations from 2003-2007 to the model.

Table 19 of Appendix H shows the marginal effects of the best-fitted model at different rates for the statutory taxes. Because the maximum of the taxes is 44.4% and to prevent a skewed estimation, four periods of coefficients of the marginal effects are selected.[12] Following Table 19, the results are similar to the results of Table 4 in Section 6.3. In conclusion, using the statutory tax rates does not significantly change the model as similar effects to the main results are found. The diminishing (significant) marginal effects for the statutory taxes and increasing marginal effects for the potential accessibility by air that are significant are high levels of taxes present evidence for a tradeoff in the FDI determinants. In line with Janský (2019), the effects are less extreme than for the EATR rates, as the differences between countries and companies are larger for the EATR. As a result, the robustness estimations of the statutory taxes improve the generalizability and reliability of the results of the main model.

7.2 Fixed Effects Poisson Model

Allison and Waterman (2002) state that the unconditional negative binomial regression and the fixed effect Poisson model should estimate relatively similar results. The model choice relies on the findings of their simulation study as they find no evidence for the IPP. However, a counterargument is that the favorable outcome may be specific to their considered simulation design and that overdispersion should

[10] Note that fewer regions meet the restrictions for a knowledge specialization. This increases the effect of the marginal effects as these region with a relatively high number of attracted investments.

[11] Model specification (3) is not further considered as adding the change of GDP significantly decreases the goodness-of-fit parameter of the adjusted McFadden's R^2 and increases the AIC and BIC values.

[12] This is selected at the percentages levels of 10, 20, 30 and 40%.

only be a serious issue unless the aim is to compute probabilities of particular counts. If so, the fixed effects Poisson model could provide more robust estimates (Cameron & Trivedi, 1998; Wooldridge, 1999).

To consider both models and challenge the unconditional NBRM model, this section adds the fixed effects Poisson models based on two models. First, the unrestricted model with interaction terms of the main results will be considered in a fixed effects Poisson setting. Second, the unrestricted interaction model of Section 7.2 is challenged. Tables 22 and 23 of Appendix H present the marginal effects at different rates of the EATR and statutory tax rates, similar to Sections 6.3 and 7.1. No significant effects of the taxes are presented for the fixed effects Poisson model using the EATR. Furthermore, there are no consecutive significant estimates of the FDI determinants as the tax rate increases. Therefore, this model presents no evidence of a tradeoff effect.

For the Poisson model with statutory taxes, a diminishing effect of the taxes is presented as the rates increase. Furthermore, larger (negative) marginal effects of the knowledge specialization threshold dummy are presented at increasing tax rates. This finding is not in line with previous NBRM models. However, it does show signs of a tradeoff effect between the statutory taxes and the FDI determinant of the knowledge specialization following a certain threshold. Together, this suggests that the main results regarding the corporate tax rates are stable but that small differences in the NBRM and fixed effect PRM change the findings on the potential tradeoff between FDI determinants. Differences in the model estimates are caused by the IPP (NBRM) or overdispersion (PRM), and based on the argumentation of earlier literature, one of the two models can be preferred and estimated. Therefore, the findings of the fixed effects Poisson model do not reject the main model but raise questions about the differences between de unconditional NBRM and the fixed effects PRM.

8 Concluding Remarks

This research investigates the tradeoff in policy decisions regarding FDI determinants in attracting greenfield FDI to European metropolitan clusters. To investigate this, the underlying relationship between the determinants is highlighted from two sides. On one side, three main characteristics of metropolitan regions are identified using the OLI paradigm of Dunning (2000): the Corruption Perceptions Index, the degree of global interconnectedness, and the specialization of the main activity in a knowledge sector. On the other side, the effective average tax rate is added to test its sensitivity and expected non-linear relationship with the number of investments (Lawless et al., 2018).

The papers' approach tests the underlying relationship of the FDI determinants using data on greenfield FDI across European metropolitan areas from 2003 to 2018. The unconditional negative binomial regression is estimated following Allison and Waterman (2002), using panel dummies to capture the time-invariant effects. The model estimates that the score of the CPI index and the knowledge sector specialization positively influence the attraction of greenfield FDI towards agglomeration economies.

For the corporate taxes (captured in the effective average tax rates), the second-order polynomial is added to the model to research and highlight its non-linear effect. This model finds that the EATR is negatively associated and that the squared term is positively associated with the number of greenfield investments in a European metropolitan region. This is in line with Lawless et al. (2018), indicating that the strength of the negative effect of taxes moderates as the rate of tax increases.

The tradeoff model combines all determinants of FDI in an unrestricted model. Interaction terms are added to the model to capture and expand the understanding of the underlying relationships among the main determinants. Using this unrestricted model with interaction terms, the marginal effects of the FDI determinants are estimated at different EATR rates. This method estimates different marginal effects for increasing EATR rates to compare the changes of the determinants. Evidence of a tradeoff effect is presented between the determinants as the marginal effects of corporate taxes are decreasing, and the effects of the potential accessibility by air (one of the main determinants of global interconnectedness) increase as the EATR levels increase.

The estimation of the tradeoff model can support policy decisions of European economic regions regarding the attraction of FDI. As this paper provides insights into the effects of the region's FDI determinants as they change their corporate tax rates, policymakers (local governments) can implement this information regarding their specific characteristics to influence their optimal level of FDI determinants efficiently.

While this research provides a detailed analysis of the underlying relationship of FDI determinants regarding FDI to European metropolitan clusters, the results are bounded by several limitations. This gives rise to various potential avenues for future research. First, regarding the data, the metropolitan clusters are made based on one or multiple NUTS-3 regions that represent all agglomerations of at least 250.000 inhabitants (Eurostat, 2020, 2021). Because of this classification, clusters are already of economic significance and more likely to receive at least one investment over the period of 2003-2018. This results in less observations of '0' counts, an assumption of the distribution of count data models, and affects the skewness of this distribution. Using data from all the European NUTS-3 regions would allow the agglomeration economies to be tested against non-clustered regions and allow the use of yearly averages in cross-sectional models.

Furthermore, it needs to be highlighted how European historical and socioeconomic events of the twentieth century might affect the analysis. Significant events such as WW1, WW2, and the cold war have strongly impacted economic and business decisions over the last decade. In addition, to expand on the external validity, multiple countries in other continents could be considered, resulting in a deeper understanding and possibly larger differences in the FDI determinants.

Finally, the results have shown to be not fully robust to changes in the model specifications. Differences between the unconditional negative binomial regression model and the fixed effects Poisson model occur because of potential biases in the data. The NBRM model possibly suffers from the incidental parameters problem, while the fixed effects PRM suffers from overdispersion in the data. The current model choices between these methods rely on the different arguments of earlier literature (Cameron & Trivedi, 1998; Allison & Waterman, 2002) and have no standard answer for all situations. Since proving the differences between the econometric models is not the main focus, this is beyond the scope of this study and an interesting topic for further research.

References

Abdella, E. A., Naghavi, N., & Chan Yin Fah, B. (2018). The effect of corruption, trade openness and political stability on foreign direct investment: Empirical evidence from BRIC countries. *International Journal of Advances and Applied Sciences*, *5*(3), 32–38.

Abdioğlu, N., Biniş, M., & Arslan, M. (2016). The Effect of Corporate Tax Rate on Foreign Direct Investment: A Panel Study for OECD Countries. *Ege Academic Review*, *16*(4), 599–610.

Allison, P. D., & Waterman, R. W. (2002). Fixed-Effects Negative Binomial Regression Models. *Sociological Methodology*, *32*(1), 247–265.

Barassi, M. R., & Zhou, Y. (2012). The effect of corruption on FDI: A parametric and non-parametric analysis. *European Journal of Political Economy*, *28*(3), 302–312.

Barrios, S., Huizinga, H., Laeven, L., & Nicodème, G. (2012). International taxation and multinational firm location decisions. *Journal of Public Economics*, *96*(11–12), 946–958.

Basile, R., Castellani, D., & Zanfei, A. (2008). Location choices of multinational firms in Europe: The role of EU cohesion policy. *Journal of International Economics*, *74*(2), 328–340.

Bel, G., & Fageda, X. (2008). Getting there fast: globalization, intercontinental flights and location of headquarters. *Journal of Economic Geography*, *8*(4), 471–495.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1981). Regression Diagnostics -- Identifying Influential Data and Sources of Collinearity. *Journal of the Operational Research Society*, *32*(2), 157.

Bijsterbosch, M., & Kolasa, M. (2010). FDI and productivity convergence in Central and Eastern Europe: an industry-level investigation. *Review of World Economics*, *145*(4), 689–712.

Blanton, S. L., & Blanton, R. G. (2007). What Attracts Foreign Investors? An Examination of Human Rights and Foreign Direct Investment. *The Journal of Politics*, *69*(1), 143–155.

Brülhart, M., Jametti, M., & Schmidheiny, K. (2012). Do Agglomeration Economies Reduce the Sensitivity of Firm Location to Tax Differentials? *The Economic Journal*, *122*(563), 1069–1093.

Buis, M. L. (2010). Stata Tip 87: Interpretation of Interactions in Nonlinear Models. *Stata Journal*, *10*(2), 305–308.

Burger, M. J., Schalk, J., Schiller, D., & Stavropoulos, S. (2021). Regional Policy and Greenfield Investments in German Districts. *Urban Science*, *5*(3), 51.

Burger, M. J., Van Der Knaap, B., & Wall, R. S. (2013). Revealed competition for greenfield investments between European regions. *Journal of Economic Geography*, *13*(4), 619–648.

Cameron, A. C., & Trivedi, P. K. (1998). Regression Analysis of Count Data. *Cambridge University Press*.

Casi, L., & Resmini, L. (2014). Spatial complexity and interactions in the FDI attractiveness of regions. *Papers in Regional Science*, *93*, 51–78.

Charbonneau, K. B. (2012). Multiple Fixed Effects in Nonlinear Panel Data Models Theory and Evidence. In *Princeton University*. Princeton University.

Cook, G. a. S., Pandit, N. R., Beaverstock, J. V., Taylor, P. J., & Pain, K. (2004). The Clustering of Financial Services in London. *44th Congress of the European Regional Science Association: "Regions and Fiscal Federalism*, 1–54.

Crozet, M., Mayer, T., & Mucchielli, J. L. (2004). How do firms agglomerate? A study of FDI in France. *Regional Science and Urban Economics*, *34*(1), 27–54.

Currie, M. (2017). Could Unilever and Shell say 'vaarwel' to their London listings? *Financial Times*. https://www.ft.com/content/87941e10-d353-11e7-ae3e-563c04c5339a

Davies, R. B., Norbäck, P. J., & Tekin-Koru, A. (2009). The Effect of Tax Treaties on Multinational Firms: New Evidence from Microdata. *World Economy*, *32*(1), 77–110.

De Mooij, R. A., & Nicodème, G. (2008). Corporate tax policy and incorporation in the EU. *International Tax and Public Finance*, *15*(4), 478–498.

Devereux, M., & Griffith, R. (2003). Evaluating Tax Policy for Location Decisions. *Social Science Research Network*, 1–31.

Devereux, M. P., & Griffith, R. (2002). The Impact of Corporate Taxation on the Location of Capital: A Review. *Economic Analysis and Policy*, *33*(2), 275–292.

Dunning, J. H. (1973). The Determinants of International Production. *Oxford Economic Papers*, *25*(3), 289–336.

Dunning, J. H. (1982). International Production and the Multinational Enterprise. *The Economic Journal*, *92*(366), 434.

Dunning, J. H. (2000). The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review*, *9*(2), 163–190.

ESPON. (2023). *Search | ESPON Database Portal*. https://database.espon.eu/search/?f=type_exact:Indicator

European Commission. (2016). *Common Consolidated Corporate Tax Base (CCCTB)*. Taxation and Customs Union. https://taxation-customs.ec.europa.eu/common-consolidated-corporate-tax-base-ccctb_en

European Commission. (2019). *The Anti Tax Avoidance Directive*. Taxation and Customs Union. https://taxation-customs.ec.europa.eu/anti-tax-avoidance-directive_en

European Commission. (2022). *Data on Taxation*. Taxation and Customs Union. https://taxation-customs.ec.europa.eu/taxation-1/economic-analysis-taxation/data-taxation_en

European Commission. (2023). *AMECO database*. https://economy-finance.ec.europa.eu/economic-research-and-databases/economic-databases/ameco-database_en

European Union. (2023). *Trade*. https://european-union.europa.eu/priorities-and-actions/actions-topic/trade_en

Eurostat. (2020). *Glossary:European Economic Area (EEA) - Statistics Explained*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_Economic_Area_(EEA)

Eurostat. (2021). *Background - Metropolitan regions*. https://ec.europa.eu/eurostat/web/metropolitan-regions/background

Eurostat. (2022a). *Corruption Perceptions Index (source: Transparency International)*. https://ec.europa.eu/eurostat/databrowser/view/SDG_16_50/default/table?lang=en&category=sdg.sdg_16

Eurostat. (2022b). *Transport statistics at regional level*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Transport_statistics_at_regional_level&oldid=573656

Eurostat. (2023a). *Gross domestic product (GDP) at current market prices by metropolitan regions*. https://ec.europa.eu/eurostat/databrowser/view/MET_10R_3GDP__custom_4706739/settings_1/table?lang=en

Eurostat. (2023b). *Metropolitan Regions Database*. https://ec.europa.eu/eurostat/web/metropolitan-regions/data/database?p_p_id=NavTreeportletprod_WAR_NavTreeportletprod_INSTANCE_RqHw4tdUJ2kl&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view

Eurostat. (2023c). *National annual road freight transport by regions of unloading*. https://ec.europa.eu/eurostat/databrowser/view/met_road_gonu/default/table?lang=en

Farhadi, M. (2015). Transport infrastructure and long-run economic growth in OECD countries. *Transportation Research Part A-policy and Practice*, *74*, 73–90.

Financial Times. (2022). *fDi Markets*. The Financial Times Ltd. https://www.fdiintelligence.com/fdi-markets?utm_source=google&utm_medium=cpc&utm_campaign=10594705413&utm_content=450638705180&utm_term=fdi%20markets%20database&matchtype=b&keyword=fdi%20markets%20database&campaignname=[SRC]-fDi-Intelligence:-fDi-Markets-LG&cid=10594705413&agid=105117976552&device=c&placement=&creative=450638705180&target=&adposition=&network=g&gclid=CjwKCAjwlqOXBhBqEiwA-hhitPF1QFc4xBuLRYV-tMnz0IzU_6wBijVqLaysYE1CtMh_kIBjyxLywhoC6GIQAvD_BwE

Franca, S., Murciego, A., Pablo Salado, J., Sisti, E., & Wilson, J. (2021). European Cluster Panorama 2021. In *European Cluster Collaboration Platform*. European Union. https://clustercollaboration.eu/sites/default/files/2021-12/European_Cluster_Panorama_Report_0.pdf

Glaeser, E. L. (2010). Agglomeration Economics. *National Bureau of Economic Research*, 1–14.

Goodspeed, T., Martinez-Vazquez, J., & Zhang, L. (2007). Are Other Government Policies More Important than Taxation in Attracting FDI? *SSRN Electronic Journal*.

Gopinath, C. (2008). Recognizing and Justifying Private Corruption. *Journal of Business Ethics*, *82*(3), 747–754.

Gordon, I. (1999). Internationalisation and Urban Competition. *Urban Studies*, *36*(5–6), 1001–1016.

Guimarães, P. R. (2008). The fixed effects negative binomial model revisited. *Economics Letters*, *99*(1), 63–66.

Hakkala, K. N., Norbäck, P., & Svaleryd, H. (2008). Asymmetric Effects of Corruption on FDI: Evidence from Swedish Multinational Firms. *The Review of Economics and Statistics*, *90*(4), 627–642.

Hansen, H., & Rand, J. (2006). On the Causal Links Between FDI and Growth in Developing Countries. *The World Economy*, *29*(1), 21–41.

Heyden, Y. V., Nijhuis, A., Smeyers-Verbeke, J., Vandeginste, B. G. M., & Massart, D. (2001). Guidance for robustness/ruggedness tests in method validation. *Journal of Pharmaceutical and Biomedical Analysis*, *24*(5–6), 723–753.

Jakovljevic, M., Camilleri, C., Rancic, N., Grima, S., Jurisevic, M., Grech, K., & Buttigieg, S. C. (2018). Cold War Legacy in Public and Private Health Spending in Europe. *Frontiers in Public Health*, *6*, 215. https://doi.org/10.3389/fpubh.2018.00215

Janský, P. (2019). Effective Tax Rates of Multinational Enterprises in the EU. *WTS*. https://www.wts.com/wts.de/publications/wts-tax-weekly/anhange/2019/2019_3_1_studie.pdf

Javorcik, B. S., & Wei, S. (2009). Corruption and cross-border investment in emerging markets: Firm-level evidence. *Journal of International Money and Finance*, *28*(4), 605–624.

Karaca-Mandic, P., Norton, E. C., & Dowd, B. E. (2012). Interaction Terms in Nonlinear Models. *Health Services Research*, *47*(1), 255–274.

Kloosterman, R. C., & Lambregts, B. (2001). Clustering of Economic Activities in Polycentric Urban Regions: The Case of the Randstad. *Urban Studies*, *38*(4), 717–732.
Lall, S., & Narula, R. (2004). Foreign Direct Investment and its Role in Economic Development: Do We Need a New Agenda? *The European Journal of Development Research*, *16*(3), 447–464.
Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*(2), 391–413.
Lawless, M., McCoy, D., Morgenroth, E. L. W., & O"Toole, C. M. (2018). Corporate tax and location choice for multinational firms. *Applied Economics*, *50*(26), 2920–2931.
Lawless, M., McCoy, D., Morgenroth, E., & O'Toole, C. (2014). *The Importance of Corporation Tax Policy in the Location Choices of Multinational Firms*. An Roinn Airgeadais Department of Finance.
Long, J. S., & Freese, J. (2014). *Regression Models for Categorical Dependent Variables Using Stata, Third Edition*. Stata Press.
Mansfield, E. D., & Helms, B. P. (1982). Detecting Multicollinearity. *The American Statistician*, *36*(3a), 158–160.
McCann, P., & Shefer, D. (2004). Location, agglomeration and infrastructure. *Papers in Regional Science*, *83*(1), 177–196.
Mehic, E., Silajdzic, S., & Babic-Hodovic, V. (2013). The Impact of FDI on Economic Growth: Some Evidence from Southeast Europe. *Emerging Markets Finance and Trade*, *49*, 5–20.
Morrisset, J., & Pirnia, N. (1999). How tax policy and incentives affect foreign direct investment - a review. *Research Papers in Economics*.
NCSS. (2023). *Negative Binomial Regression*. NCSS Statistical Software. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative_Binomial_Regression.pdf
OECD. (2019). *FDI Qualities Indicators: Measuring the sustainable development impacts of investment*. OECD. https://www.oecd.org/investment/FDI-Qualities-Indicators-Highlights.pdf
OECD. (2021). Corporate Tax Statistics - Third Edition. In *OECD*. https://www.oecd.org/tax/tax-policy/corporate-tax-statistics-third-edition.pdf
OECD. (2022a). *Corporate Tax Statistics: Fourth Edition - OECD*. https://www.oecd.org/tax/tax-policy/corporate-tax-statistics-database.htm
OECD. (2022b). *Effective Tax Rates*. https://stats.oecd.org/index.aspx?DataSetCode=CTS_ETR
ParlGov. (2023). *ParlGov Project*. https://www.parlgov.org/
Pitelis, C. N. (2009). The Sustainable Competitive Advantage and Catching-up of Nations: FDI, Clusters and the Liability (Asset) of Smallness. *Management International Review*, *49*(1), 95–120.
Port of Rotterdam. (2021). Improving the Port of Rotterdam Hinterland Accessibility and Transport Supply Chains. In *Port of Rotterdam*. Port of Rotterdam Authority. https://www.portofrotterdam.com/sites/default/files/2021-05/position-paper-eu-improving-the-hinterland-accessibility-and-transport-supply-chains_0.pdf
Porter, M. E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, 77–90.
Rasciute, S., & Downward, P. (2017). Explaining variability in the investment location choices of MNEs: An exploration of country, industry and firm effects. *International Business Review*, *26*(4), 605–613.
Robinson, C. H., & Schumacker, R. E. (2009). Interaction Effects: Centering, Variance Inflation Factor, and Interpretation Issues. *Multiple Linear Regression Viewpoints*, *35*(1), 6–11.
Rugman, A. M. (2010). Reconciling Internalization Theory and the Eclectic Paradigm. *Multinational Business Review*, *18*(2), 1–12.
Scott, A. J. (1992). The Roepke Lecture in Economic Geography the Collective Order of Flexible Production Agglomerations: Lessons for Local Economic Development Policy and Strategic Choice. *Economic Geography*, *68*(3), 219.
Spiekermann, K., & Wegener, M. (1996). Trans-European networks and unequal accessibility in Europe. *European Journal of Regional Development*, *4*(96), 35–42.
Spiekermann, K., & Wegener, M. (2006). Accessibility and spatial Development in Europe. *Scienze Regionali*, *5*(2).
Spies, J. (2010). Network and border effects: Where do foreign multinationals locate in Germany? *Regional Science and Urban Economics*, *40*(1), 20–32.
StataCorp. (2017). *mepoisson*. Stata. https://www.stata.com/manuals/memepoisson.pdf
StataCorp. (2021). *STATA Longitudinal - Data/Panel-Data Reference Manual Release 17*. StataCorp LLC. https://www.stata.com/manuals/xt.pdf
Statista. (2021). *Corporate income tax rates in Europe in 2021*. https://www.statista.com/statistics/1253236/combined-income-tax-rate-europe/
Statistics Canada. (2022). *Statistics Canada: Canada's national statistical agency*. https://www23.statcan.gc.ca/imdb-bmdi/document/2514_D2_T9_V1-eng.pdf
The World Bank. (2023). *Inflation, consumer prices*. https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=EU

Transparency International. (2012). *Corruption Perceptions Index 2003-2012*. Transparency.org. https://www.transparency.org/en/cpi/2012

UNCTAD. (2019). *Chapter 1 - Global trends and prospects*. World Investment Report. https://worldinvestmentreport.unctad.org/world-investment-report-2019/ch1-global-trends-and-prospects/

UNCTAD. (2022). *Chapter 1 - Global Investment Trends and Prospects*. World Investment Report 2022. https://worldinvestmentreport.unctad.org/world-investment-report-2022/chapter-1---global-investment-trends-and-prospects/

University of Virginia. (2020). *Understanding Robust Standard Errors*. https://data.library.virginia.edu/understanding-robust-standard-errors/

Vijayakumar, N., Sridharan, P., & Rao, K. C. S. (2010). Determinants of FDI in BRICS Countries: A panel analysis. *International Journal of Business*, *5*(3), 1–13.

White, G. C., & Bennetts, R. E. (1996). Analysis of Frequency Count Data Using the Negative Binomial Distribution. *Ecology*, *77*(8), 2549–2557.

Wijeweera, A., Dollery, B., & Clark, D. (2007). Corporate tax rates and foreign direct investment in the United States. *Applied Economics*, *39*(1), 109–117.

Williams, R. (2012). Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects. *Stata Journal*, *12*(2), 308–331.

Williams, R. (2015). *Multicollinearity*. Richard Williams. https://www3.nd.edu/~rwilliam/stats2/l11.pdf

Williams, R. (2021). *Models for Count Outcomes*. Richard Williams. https://www3.nd.edu/~rwilliam/stats3/CountModels.pdf

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, *90*(1), 77–97.

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.

World Bank Group. (2022). World Bank. https://www.worldbank.org/en/home

World Economic Forum. (2013). *The Travel & Tourism Competitiveness Report 2013*. https://www3.weforum.org/docs/TTCR/2013/TTCR_DataTables7_2013.pdf

Yaacob, W. Z. W., Lazim, M. a. S. M., & Wah, Y. B. (2011). Applying Fixed Effects Panel Count Model to Examine Road Accident Occurrence. *Journal of Applied Sciences*, *11*(7), 1185–1191.

Appendix

A. Specialization of Main Sector Activity

Table 5: List of Knowledge Intensive Sectors

Knowledge Intensive Sector	NAICS Code	Knowledge Intensive Sector	NAICS Code
Aerospace	33-	Financial Services	52-
Alternative/Renewable energy	33-	Healthcare	52- / 62-
Biotechnology	54-	Medical Devices	33- / 62-
Business Services	52- / 55-	Pharmaceuticals	32- / 62-
Chemicals	32-	Semiconductors	32- / 33-
Communications	51- / 56-	Software & IT services	51- / 81-
Consumer Electronics	53-	Space & Defense	33- / 92-

NOTES: This table visualizes the knowledge intensive sectors based on the NAICS codes as presented in Statistics Canada (2022). NAICS codes are based on the two-digit codes to give a broad overview of the knowledge sectors.

B. Descriptive Statistics

Table 6: Descriptive Statistics

Description	Obs.	Mean	St. dev.	Min	Max
Number of investments per region	4,080	8.178	19.880	0	352
Corruption Perceptions Index (CPI)	4,079	68.032	15.927	28	97
Potential multimodal accessibility (ln)	4,016	4.546	0.406	3.197	5.224
Potential accessibility by road (ln)	4,016	4.410	0.906	-0.174	5.494
Potential accessibility by rail (ln)	4,016	4.332	1.029	-0.71	5.421
Potential accessibility by air (ln)	4,016	4.531	0.402	3.47	5.222
Knowledge sector specialization	4,080	0.297	0.457	0	1
Effective Average Tax Rate (EATR) (%)	2,684	25.690	6.613	8.800	38.400
Local GDP (ln)	3,974	9.819	0.973	6.871	13.500
Population per km^2 (ln)	3,514	5.475	0.876	3.638	8.336
Unemployment (%)	3,228	9.062	5.626	0.700	42.300
Statutory taxes (%)	3,968	28.783	7.207	10.000	44.400
Compensation per employee (ln)	4,080	4.532	0.575	2.342	5.484
Inflation (%)	4,080	1.823	1.719	-4.48	15.40
EATR high	2,684	0.216	0.412	0	1
EATR medium	2,684	0.543	0.498	0	1
EATR low	2,684	0.241	0.428	0	1
Statutory taxes high (%)	3,968	0.381	0.486	0	1
Statutory taxes medium (%)	3,968	0.394	0.489	0	1
Statutory taxes low (%)	3,968	0.186	0.394	0	1
Knowledge specialization threshold	4,080	0.055	0.228	0	1
Change in GDP (%)	3,964	3.388	5.409	-24.226	46.712

NOTES: This table represents the descriptive statistics of the main and control variables over the period of 2003-2018 as introduced in Section 4.2. Missing observations are caused by missing years or incomplete region data.

Sources:

fDi Markets (2022): Number of greenfield investments per region, knowledge sector specialization, knowledge specialization threshold

ESPON database (2023): Potential accessibility indicators

Eurostat (2022a, 2023a, 2023b, 2023c): CPI Score, local GDP, population per km^2 and unemployment, change in GDP

European Commission (2022, 2023): Potential accessibility, EATR, statutory taxes & compensation per employee

The World Bank (2023): Inflation

C. Correlation Tables

Table 7 Correlations Between Explanatory Variables

		1.	2.	3.	4.	5.	6.	7.
1.	Corruption Perceptions Index (CPI)	1.000						
2.	Potential multimodal accessibility (ln)	0.447***	1.000					
3.	Potential accessibility by road (ln)	0.379***	0.764***	1.000				
4.	Potential accessibility by rail (ln)	0.377***	0.735***	0.963***	1.000			
5.	Potential accessibility by air (ln)	0.334***	0.926***	0.549***	0.520***	1.000		
6.	Knowledge sector specialization	0.119***	0.102***	-0.037**	-0.024	0.152***	1.000	
7.	Effective Average Tax Rate (EATR)	0.376***	0.275***	0.361***	0.368***	0.116***	-0.072***	1.000
8.	Local GDP (ln)	0.319***	0.438***	0.183***	0.196***	0.482***	0.362***	0.304***
9.	Population per km^2 (ln)	0.119***	0.485***	0.335***	0.278***	0.462***	0.153***	0.069***
10.	Unemployment	-0.311***	-0.374***	-0.363***	-0.336***	-0.309***	-0.051***	0.226***
11.	Statutory taxes	0.395***	0.394***	0.433***	0.415***	0.264***	-0.039**	0.919***
12.	Compensation per employee (ln)	0.714***	0.477***	0.394***	0.397***	0.367***	0.093***	0.728***
13.	Inflation	-0.327***	-0.265***	-0.252***	-0.238***	-0.217***	-0.023	-0.242***

NOTES: this table presents the correlations between the main explanatory variables and the set of control variables which are included in the models of Section 6. *P < 0.1; **P < 0.05; ***P < 0.01.

Table 8: Correlations Between Explanatory Variables (continued)

		8.	9.	10.	11.	12.	13.
1.	Corruption Perceptions Index (CPI)						
2.	Potential multimodal accessibility (ln)						
3.	Potential accessibility by road (ln)						
4.	Potential accessibility by rail (ln)						
5.	Potential accessibility by air (ln)						
6.	Knowledge sector specialization						
7.	Effective Average Tax Rate (EATR)						
8.	Local GDP (ln)	1.000					
9.	Population per km^2 (ln)	0.459***	1.000				
10.	Unemployment	-0.075***	-0.050***	1.000			
11.	Statutory taxes	0.282***	0.187***	0.047***	1.000		
12.	Compensation per employee (ln)	0.481***	0.178***	-0.026	0.711***	1.000	
13.	Inflation	-0.260***	-0.105***	-0.096***	-0.226***	-0.353***	1.000

NOTES: this table presents the correlations between the main explanatory variables and the set of control variables which are included in the models of Section 6. *P < 0.1; **P < 0.05; ***P < 0.01.

D. Count Data Model Selection Criteria

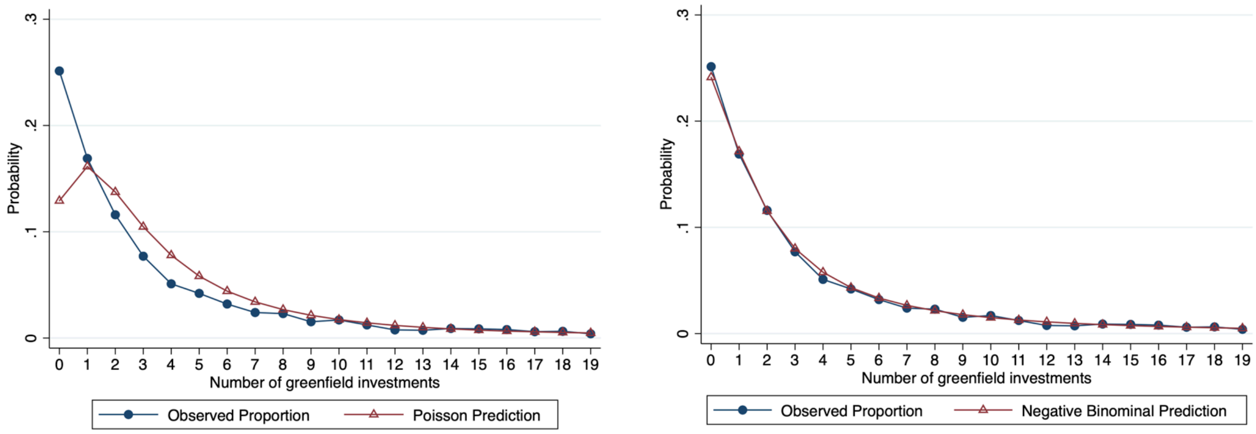

Figure 4: Model Fit for the Poisson model and Negative Binomial Model

Notes: This figure plots the observed and predicted numbers of greenfield investments in European metropolitan regions using the Poisson model and the negative binomial regression. This figure uses the third model specification from Table 10.

Table 9: Comparison of Unconditional Negative Binomial Regression Model and Fixed Effects Poisson Model

	Unconditional Negative Binomial Regression Model	Fixed Effects Poisson Model
Main variables		
CPI Score	-0.004	-0.007*
	(0.002)	(0.004)
Potential accessibility by road (ln)	-0.114**	-0.174*
	(0.058)	(0.104)
Potential accessibility by air (ln)	0.103	0.102
	(0.075)	(0.128)
Knowledge sector specialization	0.117***	0.034
	(0.027)	(0.050)

Notes: This table presents estimation results of the unconditional negative binomial model and the fixed effects Poisson model. To show that there are similarities in the coefficients of these models, the main variables of model (3) from Table x are included. The panel dummies are not added in the table as they are only included in the NBRM. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

E. NBRM Selection Estimators

The negative binomial regression model presents multiple goodness-of-fit statistics that can be used in the model selection of the specification with the best fit. The model presents McFadden's R^2, the adjusted McFadden's R^2 the Akaike (AIC) and the Bayesian (BIC) information statistics (White & Bennetts, 1996). Furthermore, tests can be performed on the overdispersion parameter α, using H_0: $\alpha = 0$, since the NBRM reduces to the Poisson model when α = 0. If the tests on α are significant, the Poisson predicts less zeros than observed zeroes and that the NBRM method fits the data better (sign of overdispersion). This is tested by LR test of $\alpha = 0$, as presented next to the goodness-of-fit statistics in the NBRM models (Long & Freese, 2014; Williams, 2021).[13]

Table 10 presents the results for the different model specifications of Hypothesis 1. Model (1) presents the control variables as a check of their effect, sign, and significance. Hereafter, the main variables are added in the model specifications. (2) presents the model including the multimodal transport. (3) and (4) present the separate potential accessibility variables included next to air with road and rail, respectively. Based on the goodness-of-fit statistics in the bottom part of Table 10, specification (3) yields the best model fit. This results from the highest McFadden's R^2 and the lowest AIC and BIC statistics. Therefore, the focus of Hypothesis 1 is directed towards this model specification for the empirical analysis.

For the second hypothesis, the non-linearity of taxes will be added and tested in different model specifications. This is presented in Table 12. Model (1) presents the negative binomial model with the EATR as main variable. Model (2) adds the high and medium dummies for the EATR (the remaining observations are low and this variable is therefore omitted in the model). Although both variables are negatively and significantly related to the number of investments, an increase in the standard error and variance is noticeable, negatively impacting the validity of the estimation. In model specification (3), the EATR is added together with its squared term, and model (4) is similar by adding the third-degree polynomial. The goodness-of-fit statistics show a slight preference for model (2) and a lot of similarities between model (3) and (4). However, because the goodness-of-fit parameters are very close to each other, the model choice arguments are based on non-linearity estimation criteria. First, model (2) is dropped because the standard errors are increasing significantly. Regarding model (4), the third-degree polynomial affects the significance of the squared EATR term, is hard to fully interpret, and again has very similar goodness of fit parameters as the other models. Furthermore, to find the best model fit regarding the non-linearity of taxes, the residuals will be plotted to visualize the functional form with randomly distributed residuals (Figure 5). Comparing the last two model specifications of Table 4, it is concluded that adding more polynomials in the functional form does not improve the model significantly

---

[13] Wald test of ln(α) corresponds to testing H_0: $\alpha = 1$ and is therefore not right for predicting the model fit.

and generates no visible changes in the residual spread. Therefore, model (3) is the preferred model that yields the best model fit, and the focus will be shifted towards this specification.

The third hypothesis is the combination of the best models fits from Hypothesis 1 and 2. Therefore, model specification (3) of Table 10-11 and specification (3) of Table 12-13 are combined to a unrestricted model including all variables. This model is further discussed in Section 6.3.

Table 10: NBRM for FDI determinants on the Number of Greenfield Investments

	(1)	(2)	(3)	(4)
Main variables				
CPI Score		0.011***	0.011***	0.011***
		(0.002)	(0.002)	(0.002)
Potential multimodal accessibility (ln)		-0.076		
		(0.086)		
Potential accessibility by road (ln)			-0.178***	
			(0.039)	
Potential accessibility by rail (ln)				-0.058***
				(0.033)
Potential accessibility by air (ln)			0.148*	0.054***
			(0.079)	(0.080)
Knowledge sector specialization		0.461***	0.436***	0.457***
		(0.042)	(0.042)	(0.042)
Control variables				
Local GDP (ln)	1.410***	1.306***	1.249***	1.284***
	(0.029)	(0.029)	(0.031)	(0.031)
Population per km^2 (ln)	0.036	0.172**	0.107***	0.076**
	(0.030)	(0.033)	(0.033)	(0.032)
Unemployment (%)	-0.008**	0.001	-0.006	-0.000
	(0.004)	(0.005)	(0.005)	(0.005)
Compensation per employee (ln)	-1.604***	-1.756***	-1.658***	-0.731***
	(0.046)	(0.062)	(0.063)	(0.062)
Inflation (%)	0.109***	0.125***	0.117***	0.123***
	(0.016)	(0.018)	(0.018)	(0.018)
Constant	-5.970***	-5.129***	-5.312***	-5.359***
	(0.320)	(0.395)	(0.380)	(0.386)
Observations	3,001	3,000	3,000	3,000
LR test of α	10,000***	9,294***	8,674***	9,072***
McFadden's R^2	0.166	0.175	0.177	0.175
Adjusted McFadden's R^2	0.164	0.172	0.174	0.172
AIC	14,902	14,748	14,720	14,747
BIC	15,034	14,898	14,876	14,903

Notes: This table presents estimation results of the negative binomial model for four model specifications. Model (1) presents the baseline model using only control variables, Model (2) includes main variables using the potential accessibility on multimodal transport. (3) and (4) present the model separately adding the main variables of road and rail, respectively. The coefficients of the added yearly dummies to capture the time invariant effects are presented in Table 11. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 11: NBRM for FDI determinants on the Number of Greenfield Investments (Continued)

	(1)	(2)	(3)	(4)
Panel Dummies				
2004	0.111	0.092	0.095	0.095
	(0.169)	(0.178)	(0.175)	(0.178)
2005	0.223	0.178	0.180	0.182
	(0.158)	(0.165)	(0.161)	(0.164)
2006	0.330**	0.259*	0.272*	0.265*
	(0.141)	(0.148)	(0.146)	(0.147)
2007	0.037	0.017	0.032	0.025
	(0.138)	(0.144)	(0.143)	(0.144)
2008	0.039	0.002	0.003	0.016
	(0.132)	(0.139)	(0.138)	(0.139)
2009	0.182	0.162	0.161	0.167
	(0.149)	(0.157)	(0.154)	(0.156)
2010	0.071	-0.014	0.004	-0.006
	(0.149)	(0.153)	(0.150)	(0.152)
2011	0.031	-0.024	-0.008	-0.021
	(0.142)	(0.147)	(0.144)	(0.146)
2012	0.219	0.125	0.140	0.127
	(0.149)	(0.151)	(0.147)	(0.150)
2013	0.936***	0.901***	0.909***	0.905***
	(0.151)	(0.154)	(0.151)	(0.154)
2014	0.437***	0.474***	0.463***	0.470***
	(0.164)	(0.172)	(0.167)	(0.170)
2015	0.462***	0.488***	0.476***	0.485***
	(0.173)	(0.182)	(0.178)	(0.181)
2016	0.269*	0.330**	0.333**	0.333**
	(0.151)	(0.160)	(0.157)	(0.160)
2017	0.002	-0.014	0.004	-0.006
	(0.145)	(0.151)	(0.148)	(0.150)
2018	-0.026	-0.011	0.007	-0.002
	(0.144)	(0.152)	(0.150)	(0.152)

Notes: This table presents estimation results of the negative binomial model for four model specifications. Model (1) presents the baseline model using only control variables, Model (2) includes main variables using the potential accessibility on multimodal transport. (3) and (4) present the model separately adding the main variables of road and rail, respectively. The coefficients of the main and control variables are presented in Table 10. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 12: NBRM for the EATR on the Number of Greenfield Investments

	(1)	(2)	(3)	(4)
Main variables				
EATR	-0.050***		-0.207***	-0.040
	(0.005)		(0.003)	(0.082)
EATR high		-1.165***		
		(0.098)		
EATR medium		-1.078***		
		(0.088)		
$EATR^2$			0.003***	-0.004
			(0.000)	(0.003)
$EATR^3$				0.000**
				(0.000)
Control variables				
Local GDP (ln)	1.396***	1.374***	1.369***	1.367***
	(0.029)	(0.029)	(0.029)	(0.029)
Population per km^2 (ln)	0.100***	0.159***	0.162***	0.165***
	(0.031)	(0.031)	(0.033)	(0.033)
Unemployment (%)	0.005	0.003	0.005	0.006
	(0.004)	(0.004)	(0.004)	(0.004)
Compensation per employee (ln)	-0.107***	-0.850***	-0.944***	-0.946***
	(0.057)	(0.065)	(0.061)	(0.061)
Inflation (%)	0.063***	0.073***	0.059***	0.055***
	(0.020)	(0.019)	(0.020)	(0.019)
Constant	-6.932***	-8.685***	-5.836***	-6.997***
	(0.276)	(0.312)	(0.336)	(0.616)
Observations	2,133	2,133	2,133	2,133
LR test of α	6,862***	6,759***	6,736***	6,717***
McFadden's R^2	0.179	0.184	0.182	0.182
Adjusted McFadden's R^2	0.176	0.181	0.179	0.179
AIC	10,890	10,834	10,865	10,863
BIC	11,002	10,942	10,973	10,976

Notes: This table presents estimation results of the negative binomial model for four model specifications. Model (1) presents the model with only the EATR as main variable, Model (2) includes dummies for the categories of high – medium – and low EATR. (3) and (4) present the functional form of the model with the squared term and third order respectively. The coefficients of the added yearly dummies to capture the time invariant effects are presented in Table 13. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 13: NBRM for EATR on the Number of Greenfield Investments (Continued)

	(1)	(2)	(3)	(4)
Panel Dummies				
2009	-0.098	-0.056	-0.106	-0.126
	(0.116)	(0.110)	(0.113)	(0.114)
2010	-0.202*	-0.162	-0.191*	-0.206*
	(0.112)	(0.108)	(0.112)	(0.112)
2011	-0.221**	-0.171*	-0.191*	-0.202*
	(0.108)	(0.101)	(0.105)	(0.106)
2012	-0.002	0.041	0.013	-0.004
	(0.119)	(0.113)	(0.115)	(0.115)
2013	0.627***	0.675***	0.621***	0.599***
	(0.115)	(0.110)	(0.112)	(0.113)
2014	0.089	0.077	0.036	-0.019
	(0.134)	(0.126)	(0.131)	(0.136)
2015	0.133	0.116	0.069	0.009
	(0.146)	(0.135)	(0.143)	(0.146)
2016	-0.026	-0.016	-0.045	-0.090
	(0.127)	(0.119)	(0.123)	(0.125)
2017	-0.266**	-0.230**	-0.254**	-0.269**
	(0.112)	(0.105)	(0.108)	(0.108)
2018	-0.290***	-0.245***	-0.269**	-0.280***
	(0.113)	(0.107)	(0.110)	(0.111)

Notes: This table presents estimation results of the negative binomial model for four model specifications. Model (1) presents the model with only the EATR as main variable, Model (2) includes dummies for the categories of high – medium – and low EATR. (3) and (4) present the functional form of the model with the squared term and third order respectively. The coefficients of main and control variables are presented in Table 12. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

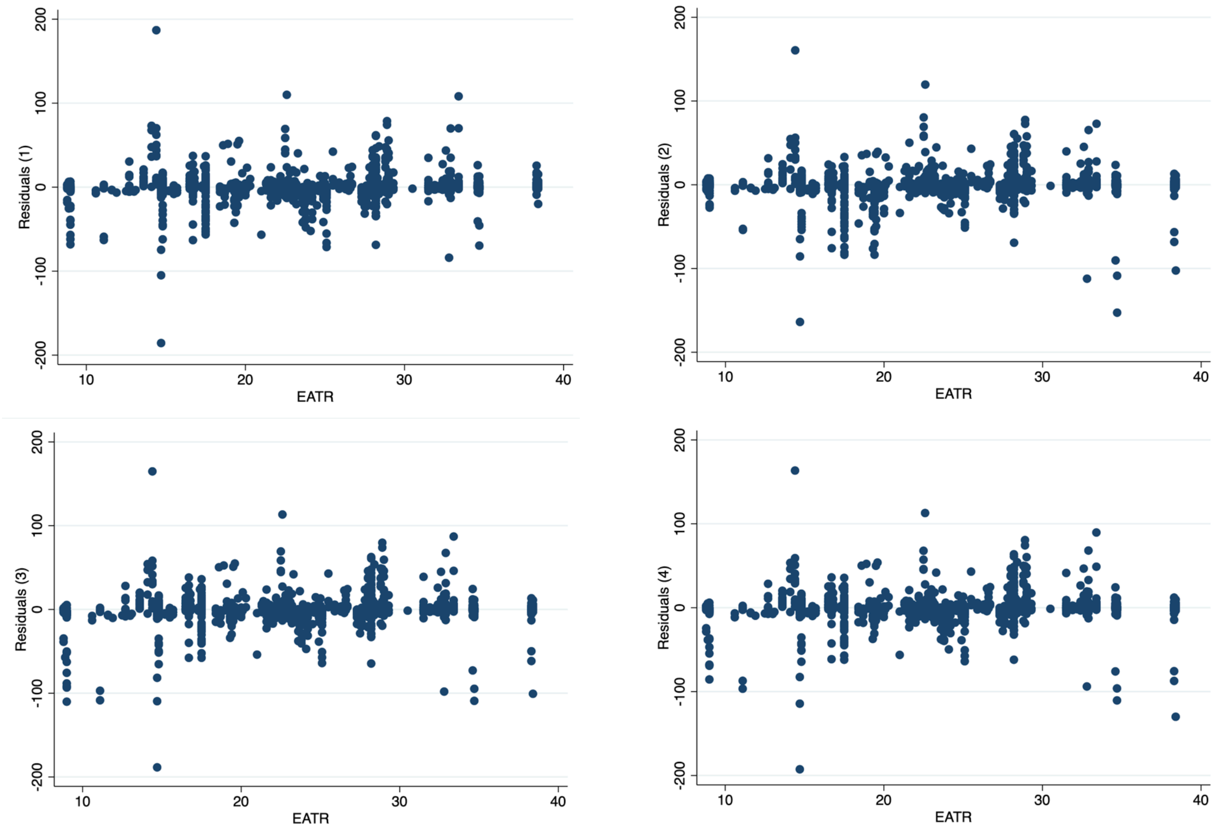

Figure 5: Residual plots for the various methods using EATR as main variable

Notes: This figure plots the residuals for the four model specifications of Table 12. Models (1) – (4) are in this case corresponding to Residuals (1) – (4) on the y-axis.

F. Unrestricted Model Estimations

Table 14: NBRM for the Unrestricted Models on the Number of Greenfield Investments

	(1)	(2)	(3)	(4)
Main variables				
CPI Score	0.021***	0.230***	0.029***	0.222***
	(0.002)	(0.025)	(0.011)	(0.025)
Potential accessibility by road (ln)	-0.008	-0.084	-0.181	-0.198
	(0.037)	(0.403)	(0.147)	(0.432)
Potential accessibility by air (ln)	0.162*	1.755*	-0.244	2.109**
	(0.090)	(0.977)	(0.216)	(0.963)
Knowledge sector specialization	0.317***	3.425***	0.047	3.101
	(0.044)	(0.480)	(0.165)	(0.465)
EATR	-0.206***	-0.615***	-0.288***	-0.452***
	(0.028)	(0.081)	(0.043)	(0.079)
$EATR^2$	0.003***		0.003***	
	(0.001)		(0.001)	
Interaction terms				
CPI Score x EATR			-0.000	
			(0.000)	
Potential accessibility by road x EATR			0.007	
			(0.006)	
Potential accessibility by air x EATR			0.018**	
			(0.008)	
Knowledge sector specialization x EATR			0.010*	
			(0.006)	
Observations	2,133		2,133	
LR test of α	5,851***		5,617***	
McFadden's R^2	0.193		0.194	
Adjusted McFadden's R^2	0.190		0.190	
AIC	10,722		10,718	
BIC	10,852		10,870	

Notes: This table presents main estimation results of the unrestricted negative binomial model for four model specifications. Model (1) presents the coefficients of the unrestricted model without interaction terms. Model (2) presents its average marginal effects. (3) and (4) present the unrestricted model including the interaction terms and its average marginal effects, respectively. The control variables and coefficients of the added yearly dummies to capture the time invariant effects are presented in Table 15. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 15: NBRM for the Unrestricted Models on the Number of Greenfield Investments (Continued)

	(1)	(2)	(3)	(4)
Control variables				
Local GDP (ln)	1.270***	13.736***	1.264***	13.682***
	(0.035)	(0.690)	(0.035)	(0.696)
Population per km^2 (ln)	0.169***	1.823***	0.160***	1.734***
	(0.036)	(0.401)	(0.035)	(0.397)
Unemployment (%)	0.022***	0.239***	0.023***	0.245***
	(0.005)	(0.059)	(0.006)	(0.062)
Compensation per employee (ln)	-1.316***	-14.231***	-1.311***	-14.195***
	(0.072)	(0.974)	(0.076)	(0.998)
Inflation (%)	0.096***	1.039***	0.094***	1.019***
	(0.021)	(0.226)	(0.021)	(0.233)
Constant	-5.901***		-3.791***	
	(0.388)		(0.380)	
Panel Dummies				
2009	-0.026	-0.269	-0.035	-0.352
	(0.121)	(1.239)	(0.123)	(1.258)
2010	-0.164	-1.556	-0.168	-1.605
	(0.115)	(1.133)	(0.116)	(1.150)
2011	-0.198*	-1.852*	-0.206*	-1.932*
	(0.109)	(1.1055)	(0.109)	(1.062)
2012	-0.012	-0.123	-0.023	-0.231
	(0.117)	(1.204)	(0.118)	(1.209)
2013	0.689***	10.228***	0.680***	10.094***
	(0.118)	(1.720)	(0.119)	(1.712)
2014	0.185	2.100	0.177	2.006
	(0.139)	(1.555)	(0.138)	(1.554)
2015	0.201	2.230	0.190	2.171
	(0.149)	(1.705)	(0.147)	(1.683)
2016	0.094	1.014	0.086	0.933
	(0.131)	(1.402)	(0.132)	(1.408)
2017	-0.194*	-1.818	-0.211*	-1.975*
	(0.114)	(1.121)	(0.114)	(1.127)
2018	-0.185	-1.743	-0.193	-1.816
	(0.119)	(1.153)	(0.119)	(1.163)

Notes: This table presents control estimation results of the unrestricted negative binomial model for four model specifications. Model (1) presents the coefficients of the unrestricted model without interaction terms. Model (2) presents its average marginal effects. (3) and (4) present the unrestricted model including the interaction terms and its average marginal effects, respectively. The main variables and interaction terms are presented in Table 14. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

G. Tradeoff Effects

Table 16: Marginal effects of Main Variables in a High – Medium – Low Scenario for the EATR

	High	Medium	Low
Main variables			
CPI Score	0.148***	0.252***	0.772***
	(0.030)	(0.033)	(0.274)
Potential accessibility by road (ln)	0.181	-0.521	-3.424
	(0.393)	(0.546)	(2.634)
Potential accessibility by air (ln)	2.458***	1.393	-1.844
	(0.861)	(1.066)	(4.325)
Knowledge sector specialization	2.808***	2.868***	4.523
	(0.401)	(0.639)	(3.288)
EATR	-0.032	-0.755***	-3.868***
	(0.054)	(0.131)	(1.271)

Notes: This table presents estimation results of the marginal effects of the main variables at different levels of the EATR. The marginal effects are based on model specification (3) from Table 14 of Appendix F The high scenario is estimated at an EATR level of 30%, the medium scenario at 20% and the low scenario at 10%. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

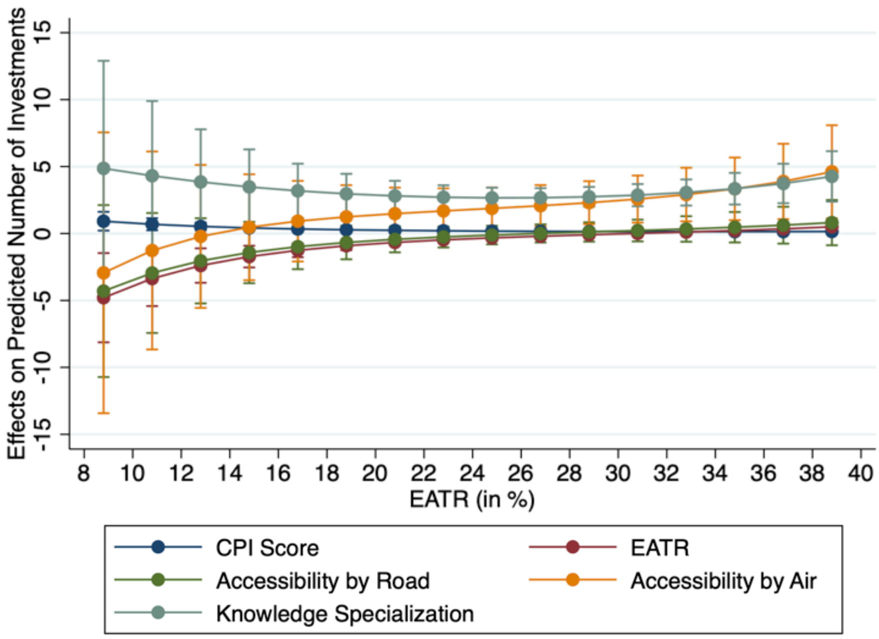

Figure 6: Marginal effects of the Unrestricted model on the Number of Investments at different EATR levels

Notes: This figure plots the marginal effects for the five main variables of Table 4 with the EATR on the x-axis and the number of greenfield investments on the y-axis. In this figure, the confidence intervals are added for each level of the EATR.

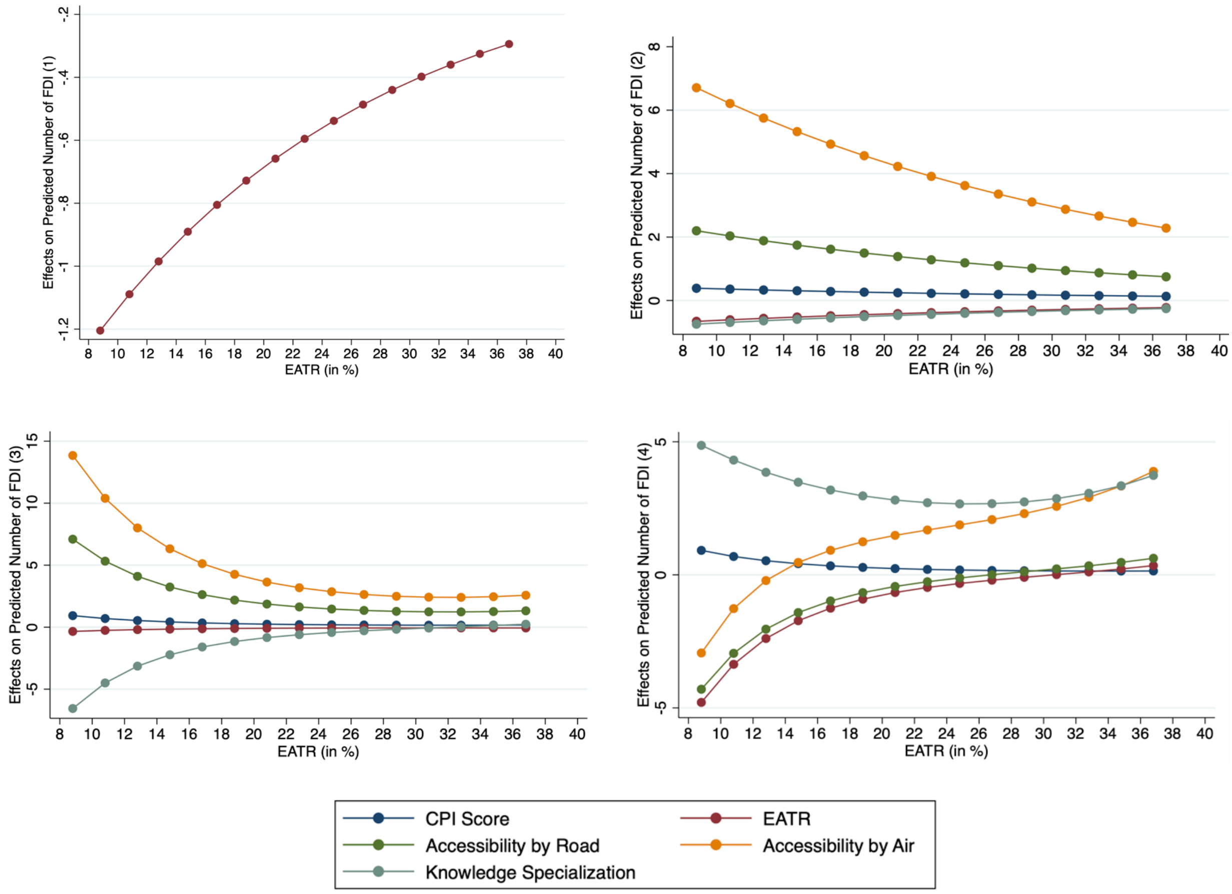

Figure 7: Marginal effects of multiple models on the Number of Investments at different EATR levels

Notes: This figure plots the marginal effects for the main variables of Table 4 with the EATR on the x-axis and the number of greenfield investments on the y-axis. Models (1) - (4) on the y-axis correspond to the restricted and unrestricted models that will be compared. Model (1) only includes the EATR, (2) includes all main variables with only the EATR. Model (3) represents the unrestricted model with all main variables and the second-order polynomial of the EATR and (4) is the unrestricted model with the interaction terms.

Addition to Figure 7:

Following from Figure 7, it is visible how the effect of taxes changes as the model restrictions differ. In restricted model (1) the effect of taxes is visible as it slowly diminishes over increasing tax rates. In model (2) the addition of the main variables present larger effects for the FDI determinants. The first unrestricted model (3) includes the second order polynomial of the EATR. This leads to diminishing marginal effects of all main determinants as they are following this squared term. In order to correctly estimate the underlying relationships of the determinants at different levels of the EATR, model (4) is estimated including interaction terms of the main variables with the EATR.

H. Outputs of Robustness Checks

Table 17: Robustness Checks for the NBRM on the Number of Greenfield Investments

	(1)	(2)	(3)	(4)
Main variables				
CPI Score	0.006***	0.008***	0.005	-0.013
	(0.002)	(0.002)	(0.003)	(0.010)
Potential accessibility by road (ln)	0.036	0.022	-0.437***	-0.302**
	(0.036)	(0.036)	(0.054)	(0.148)
Potential accessibility by air (ln)	0.093	0.080	1.511***	-0.182
	(0.073)	(0.073)	(0.086)	(0.196)
Knowledge sector specialization	0.376***			
	(0.040)			
Knowledge specialization threshold		0.525***	1.095***	0.481**
		(0.058)	(0.078)	(0.222)
Statutory tax rate	-0.066***	-0.062***	-0.119***	-0.129***
	(0.022)	(0.022)	(0.034)	(0.035)
Statutory tax rate^2	-0.000	-0.000	0.001**	-0.001**
	(0.000)	(0.000)	(0.001)	(0.001)
Interaction terms				
CPI Score x Statutory tax rate				0.001**
				(0.000)
Potential accessibility by road x Statutory tax rate				0.010**
				(0.005)
Potential accessibility by air x Statutory tax rate				0.010
				(0.007)
Knowledge sector specialization x Statutory tax rate				0.001
				(0.008)
Control variables				
Local GDP (ln)	1.248***	1.288***		1.296***
	(0.029)	(0.029)		(0.029)
Change in GDP (%)			0.029**	
			(0.006)	
Population per km^2 (ln)	0.118***	0.121***	0.696***	0.114***
	(0.031)	(0.030)	(0.038)	(0.030)
Unemployment (%)	0.008*	0.008*	0.0011*	0.006
	(0.005)	(0.005)	(0.006)	(0.005)
Compensation per employee (ln)	-1.105***	-1.173***	-0.192**	-1.103***
	(0.068)	(0.067)	(0.085)	(0.081)
Inflation (%)	0.113***	0.122***	0.049***	0.109***
	(0.016)	(0.015)	(0.019)	(0.016)

Notes: This table presents main estimation results of the unrestricted negative binomial model using statutory tax rates for four model specifications. Model (1) presents the coefficients using the statutory taxes and its second-order polynomial. Model (2) introduces the knowledge specialization threshold. Model (3) introduces the change in GDP and (4) presents the unrestricted model including the interaction terms. The coefficients of the added yearly dummies to capture the time invariant effects and the goodness-of fit parameters are presented in Table 18. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 18: Robustness Checks for the NBRM on the Number of Greenfield Investments (Continued)

	(1)	(2)	(3)	(4)
Panel Dummies				
2004	0.019	0.022	-0.093	0.017
	(0.160)	(0.158)	(0.170)	(0.158)
2005	0.025	0.047	0.138	0.018
	(0.147)	(0.145)	(0.164)	(0.145)
2006	0.131	0.170	0.298*	0.148
	(0.130)	(0.131)	(0.170)	(0.128)
2007	-0.071	-0.049	0.285	-0.056
	(0.130)	(0.131)	(0.179)	(0.128)
2008	-0.242*	-0.219	0.451**	-0.198
	(0.135)	(0.133)	(0.182)	(0.132)
2009	-0.210	-0.179	0.441**	-0.187
	(0.145)	(0.145)	(0.177)	(0.142)
2010	-0.391***	-0.317**	0.075	-0.317**
	(0.138)	(0.141)	(0.171)	(0.139)
2011	-0.423***	-0.408***	0.010	-0.396***
	(0.138)	(0.137)	(0.165)	(0.135)
2012	-0.248*	-0.199	0.330*	-0.182
	(0.142)	(0.144)	(0.173)	(0.142)
2013	0.536***	0.562***	1.092***	0.578***
	(0.139)	(0.139)	(0.168)	(0.136)
2014	0.082	0.140	0.566***	0.163
	(0.158)	(0.156)	(0.193)	(0.153)
2015	0.128	0.179	0.589***	0.193
	(0.170)	(0.168)	(0.196)	(0.164)
2016	-0.070	-0.053	0.492***	-0.050
	(0.151)	(0.151)	(0.181)	(0.148)
2017	-0.281**	-0.249*	0.410**	-0.205
	(0.140)	(0.141)	(0.183)	(0.140)
2018	-0.413***	-0.413***	0.410**	-0.398***
	(0.143)	(0.142)	(0.186)	(0.140)
Constant	-6.161***	-6.343***	-5.042***	-3.692***
	(0.334)	(0.329)	(0.439)	(0.140)
Observations	3,000	3,000	2,998	3,000
LR test of α	7,198***	7,367***	26,000***	6,997***
McFadden's R^2	0.192	0.191	0.092	0.192
Adjusted McFadden's R^2	0.189	0.188	0.089	0.188
AIC	14,450	14,474	16,216	14,463
BIC	14,618	14,643	16,384	14,655

Notes: This table presents estimation results of the unrestricted negative binomial model using statutory tax rates for four model specifications. Model (1) presents the coefficients using the statutory taxes and its second-order polynomial. Model (2) introduces the knowledge specialization threshold. Model (3) introduces the change in GAP and (4) presents the unrestricted model including the interaction terms. The coefficients of the main variables, interaction terms and control variables are presented in Table 17. The LR test statistic concerns the test of significance for overdispersion parameter α. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 19: Marginal effects at Different Rates of the Statutory Taxes

	10%	20%	30%	40%
Main variables				
CPI Score	-0.099	0.015	0.065***	0.060***
	(0.105)	(0.044)	(0.019)	(0.018)
Potential accessibility by road (ln)	-3.348**	-1.253*	0.040	0.420
	(1.447)	(0.710)	(0.309)	(0.265)
Potential accessibility by air (ln)	-1.333	0.296	0.995	0.892*
	(2.211)	(1.006)	(0.643)	(0.531)
Knowledge specialization threshold	9.930***	7.845***	4.970	2.516***
	(3.581)	(1.536)	(0.714)	(0.663)
Statutory tax rate	-0.261	-0.480***	-0.471***	-0.320***
	(0.054)	(0.130)	(0.042)	(1.271)

Notes: This table presents estimation results of the marginal effects of the main variables at different levels of the statutory taxes. The marginal effects are based on model specification (4) from Table 17 of Appendix H. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 20: Robustness Checks for the Fixed Effects Poisson Model on the Number of Greenfield FDI

	EATR	Statutory Taxes
Main variables		
CPI Score	0.062***	0.059***
	(0.022)	(0.015)
Potential accessibility by road (ln)	-0.206	0.253
	(0.283)	(0.300)
Potential accessibility by air (ln)	-0.039	-0.455
	(0.413)	(0.387)
Knowledge sector specialization	-0.214	
	(0.040)	
Knowledge specialization threshold		0.311*
		(0.189)
EATR	0.167***	
	(0.072)	
$EATR^2$	0.000	
	(0.072)	
Statutory tax rate		0.012
		(0.061)
Statutory tax $rate^2$		0.002**
		(0.001)

Notes: This table presents main estimation results of the fixed effects Poisson model using statutory tax rates for two model specifications. The first model includes the same variables as the main results, model specification (3) from Table 14-15 of Appendix F. The second model uses the statutory taxes and knowledge specialization threshold of model specification (4) from Table 17. The coefficients of the interaction terms, control variables and the goodness-of fit parameters are presented in Table 21. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 21: Robustness Checks for the Fixed Effects Poisson Model on the Number of Greenfield FDI (Continued)

	EATR	Statutory Taxes
Interaction terms		
CPI Score x Statutory tax rate	-0.003***	-0.002***
	(0.001)	(0.001)
Potential accessibility by road x Statutory tax rate	0.000	-0.017
	(0.010)	(0.011)
Potential accessibility by air x Statutory tax rate	0.001	0.017*
	(0.015)	(0.012)
Knowledge sector specialization x Statutory tax rate	0.008	-0.016**
	(0.006)	(0.008)
Control variables		
Local GDP (ln)	1.187***	1.099***
	(0.265)	(0.160)
Population per km^2 (ln)	0.914**	0.838**
	(0.386)	(0.363)
Unemployment (%)	0.042***	0.029***
	(0.012)	(0.007)
Compensation per employee (ln)	0.908***	1.169***
	(0.220)	(0.237)
Inflation (%)	0.037***	0.032***
	(0.014)	(0.012)
Observations	2,094	2,979
AIC	9,570	13,442
BIC	9,654	13,532

Notes: This table presents main estimation results of the fixed effects Poisson model using statutory tax rates for two model specifications. The first model includes the same variables as the main results, model specification (3) from Table 14-15 of Appendix F. The second model uses the statutory taxes and knowledge specialization threshold of model specification (4) from Table 17. The coefficients of the main variables are presented in Table 20. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 22: Marginal effects of the main variables at different EATR rates for the fixed effects PRM

	15%	25%	35%
Main variables			
CPI Score	0.002	-0.011	-0.023*
	(0.007)	(0.007)	(0.012)
Potential accessibility by road (ln)	-0.301**	-0.274**	-0.247
	(0.125)	(0.116)	(0.174)
Potential accessibility by air (ln)	2.245	2.215*	0.185
	(0.160)	(0.160)	(0.178)
Knowledge sector specialization	-0.012	0.007	0.028
	(0.058)	(0.042)	(0.069)
EATR	-0.007	-0.009	-0.010
	(0.025)	(0.007)	(0.013)

Notes: This table presents estimation results of the marginal effects of the main variables at different levels of the EATR. The marginal effects are based on model specification (1) from Table 20-21 of Appendix H. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.

Table 23: Marginal effects of the main variables at different statutory tax rates for the fixed effects PRM

	10%	20%	30%	40%
Main variables				
CPI Score	0.037***	0.014**	-0.008	-0.031***
	(0.010)	(0.007)	(0.006)	(0.009)
Potential accessibility by road (ln)	0.080	-0.092	-0.265**	0.437**
	(0.202)	(0.126)	(0.124)	(0.197)
Potential accessibility by air (ln)	-0.281	0.107	0.068	0.242
	(0.272)	(0.171)	(0.125)	(0.183)
Knowledge specialization threshold	0.148	-0.015***	-0.178**	-0.341***
	(0.116)	(0.054)	(0.063)	(0.129)
Statutory tax rate	-0.100***	-0.063***	-0.025***	-0.013***
	(0.026)	(0.013)	(0.011)	(0.023)

Notes: This table presents estimation results of the marginal effects of the main variables at different levels of the statutory taxes. The marginal effects are based on model specification (2) from Table 20-21 of Appendix X. *P < 0.1; **P < 0.05; ***P < 0.01. Robust standard errors are reported in parentheses.