ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis [Industrial Dynamics and Strategy]

Agglomeration Economies and Their Effect on the Location of New Ventures.

Name student: Deyan Karakachanov

Student ID number: 455452

Supervisor: Spyridon Stavropoulos

Second assessor: xxx

Date final version: 27.08.2017

Table of Contents

Summary

Agglomeration economies are argued by many economists to have a vital importance for new firms, mainly by providing them with an external source of economies of scale. This paper attempts to bring further proof that new ventures are heavily influenced by two types of agglomeration economies: localization and urbanization. The former refers to a cluster of companies from one industry, while the latter to a gathering of firms from many industries, most often in an urban environment.

Using a logit regression, I attempt to present the relationship between newly established firms and these aspects of agglomeration economies by examining the four biggest regions in the Netherlands on a Nuts 3 level. Included in the analysis are variables to account for economic and demographic conditions, road freight transport and intellectual property rights. Overall, the results are very supportive of the notion that localization economies are important for new ventures, while urbanization economies are found to have a negative and insignificant effect. Furthermore, these results were checked by examining them against regressions accounting for successful firms and across industries, which provided further support for the results.

In conclusion, the main finding of the paper is the relevant and significant effect of localization economies, as well as the surprising lack of effect of urbanization. These results are indicative of the need for more research on the topic, since no consensus has been reached in the literature.

1. Introduction

Agglomeration economies have long been proven to be an important part of the development of regions. This field relies on the basis that geographical concentration of economic agents and activities has beneficial effects for the productivity and the growth of firms within the area (Mukkala, 2004). Cities in particular are a key point in understanding agglomeration economies since they present numerous advantages to new firms. These benefits can be lower transportation costs to the final market, availability of larger and more diverse labor pool, the presence of spillovers of knowledge and expertise from the movement of labor and proximity to intermediate products, to name a few. These advantages greatly increase with a growing population, however, up to a point at which some disadvantages arise, such as congestion, increased land prices and pollution (Viladecans-Marsal, 2004). In this paper, I will discuss two aspects of agglomeration economies - localization and urbanization in the context of new ventures in the Netherlands.

The concept of localization was first proposed in Marshall (1890), where the author made a distinction between internal and external economies of scale. The first referring to the capabilities and resources of a single firm, while the second to the performance and development of the whole industry in a region. The main idea of these Marshall externalities concerns the benefits that firms in an industry receive from being clustered together in one area. If an industry has localization economies, then economic actors are likely to gather together in a small number of regions or cities, which are specialized in production within that industry and any closely interconnected activities. This leads to a full exploitation of the scale externalities without great increases in local land rent and congestion costs (Henderson, 2003). Elberts & McMillen (1999) give the example of Sillicon Valley, where many of the software firms are small. This implies that the benefits of internal economies of scale are insignificant for the majority. However, because of high risk assosiated with software startups, when some firms fail it leads to an increased pool of experiensed and mobile labor force, which in turn is advantegeous for all firms in the region. Localization economies also are present when restaurants cluster near each other in order to gain from other outlets' customer overflow and to enable comparison shopping from undecided diners (Elberts & McMillen, 1999).

In contrast, urbanization, first proposed by Jacobs (1969), refers to the economic diversity of an area. Urbanization economies can be viewed as a function of city size, because the main benefit comes from the close proximity in which multiple industries function, therefore all firms within the area receive the externalities. Consequently, a larger local environment generally equals higher benefits from the Jacobs externalities.

Urbanization economies rely on a range of place-specific and ideographic factors that are consistently urban. For some industries, these externalities compensate for the drawbacks associated with operating in a large city, such as high wages, expensive real estate and restrictive regulation. Therefore, businesses operating in high-fashion apparel and publishing manufacturing, as well as financial, business, research and development and management services tend to be clustered disproportionately in larger urban areas (Henderson, 2003).

The purpose of this paper is to examine both localization and urbanization economies in the Netherlands. The research question being examined is whether these concepts influence the choice of where to establish a new venture. The paper tackles this issue by using large dataset of Dutch firms and using only those who have been established in the years from 2011 to 2014. This approach excludes any firms who have failed in the period in question since they should not significantly affect the choice of where to establish a new firm.

The focus is on all manufacturing and service industries measured at the 2-digit level from the Standard Corporate classification (SBI) form the Central Bureau of Statistics in Netherlands. Furthermore, the analysis is conducted on the Nuts 3 level, which is generally accepted as the optimal scale for analyzing a single country. The study is concentrated in the four major regions in Netherlands – Amsterdam, Rotterdam, Utrecht and The Hague. Since the dependent variable in question is binary (whether a company is a startup or not) the method used is a logit regression, conducted over 4 years.

The paper is organized as follows: Section 2 introduces the review of relevant literature and the hypotheses I will defend in this work. Section 3 presents the data and methodology, while Section 4 summarizes the results. Section 5 presents robustness checks, Section 6 is a discussion. Section 7 concludes.

2. Literature Review

Firms operating in regions which have localization and urbanization economies benefit from the externalities stemming from working with other firms, the labor market and institutional specialization (Lorenzen & Frederiksen, 2008). In order to fully understand the concept behind these types of agglomeration economies, they will be described in relation to the way the externalities affect firms. It is vital to address information spillovers as a major agglomeration force (Fujita & Thisse, 1996). Since knowledge is a public good, usage of one firm does not exclude others from benefiting from it, therefore, these spillovers have effects very similar to externalities for the economic agents. Moreover, the advantages from communication increase with the number of agents and additionally these benefits decrease with distance. Consequently, firms have greater incentives to cluster near each other. However, the formation of clusters also brings an increase in commuting distance, wage rate and land rent. These factors generally discourage agglomeration and therefore the tradeoff between these forces brings equilibrium for firms. (Fujita & Thisse, 1996).

As stated above, localization externalities represent the benefits of a region specializing in one industry and its related activities. In the context of localization economies, the externalities received from other firms can be divided into static and dynamic externalities and competition. The static externalities represent the benefits of firms coordinating among themselves. Because of the closely related products and knowledge base, this gives great advantages for firms, such as networks, value chains, collaborations (Marshall, 1890; Lorenzen & Frederiksen, 2008). On the other hand, dynamic externalities refer to the knowledge spillovers, which can be the study and imitation that are common practice among firms. Competition in this context is beneficial because it raises productivity, as described in Porter (2000) and Lorenzen & Frederiksen (2008).

The labour market externalities also play a vital role. A specialized region generally produces a very skilled and large base of workers, which moves among firms, further raising the employees knowledge and skills base. However, this generally results in increased wages as well (Lorenzen & Frederiksen, 2008; Mukkala, 2004).

Lastly, institutional specialization can be separated into formal and informal. The benefits of formal institutions are the focus of public services and development agencies in the specific industry in which firms oprerate, resulting in a further increased knowledge base of workers in the field. Informal institutions, such as languages, cultural environment and religion can also be valuable to firms, since it

leads to decreased barries to doing business, such as transaction and time costs, and to an increased ease of knowledge spillovers (Lorenzen & Frederiksen, 2008).

Urbanization economies stem from the diversity of the firms within a region. From the variety of the industries companies receive both dynamic and static externalities. The former refers to the collaboration of different sectors, while dynamic externalities i.e. the knowledge spillovers from multiple industries, both can often lead to new and improved products (Lorenzen & Frederiksen, 2008). The main difference between localization and urbanization in this context is that the former fosters incremental improvements, while the latter can lead to a more radical innovation, because many new ideas are appropriated from different industries. Furthermore, as mentioned in Lorenzen & Frederiksen (2008), there are the venture capital investments from thriving sectors to ones in need of capital, as well as the opportunities of cheap real estate from declining industries in the region.

Urbanization economies generally bring a very varied and broad workforce, stemming from overlapping labor markets and inflow of workers, which could lead to a greater corporate and startup entrepreneurship (Jacobs, 1969; Lorenzen & Frederiksen, 2008).

The diversity in institutions provides very different advantages to firms compared to localization economies. First, most major universities and higher education institutions have always been a vital source of new talent for companies. Furthermore, when a big multinational university is present in the region it could lead to a large variety of skills and knowledge being available to employers (Lorenzen & Frederiksen, 2008). This is the case for all four of the regions in question in this paper. Second, regions which exibit urbanization externalities usually have a good level of international interconnectedness i.e. airports, marine ports, which foster the creation of clusters by making it easier to move goods, labour globally. Third, large metropolitan cities/regions provide diverse housing, general tolerance and cultural acceptance and therefore are very succesful in attracting international talent, which can be beneficial for companies.

There are also cases of both localization and urbanization economies being present in one city/region. This is possible when within a large and diverse urban area, clusters of firms benefit from both the Jacobs externalities and from their colocation (Marshall externalities). In their paper, Lorenzen and Frederiksen (2008) offer two ways in which both types of externalities interact. First, big cities offer a flow of capital, labour and knowledge to specialized clusters within them. In this case, the quality of institutions in the urban area are vital in transferring these advantages to the clusters. Although this is most relevant for extramely large megapolises, it can also happen in medium sized cities, such as the four studies in this

paper, because they are large enough to be able to accommodate both the urban environment and the clusters.

Second, the more specialized a cluster, the higher potential benefit it has from coordinating with another specialized cluster. Holm and Pedersen (2000) present the case of centres of excellence, which are clusters with a high technological understanding of their field. These clusters are very likely to foster knowledge spillovers leading to valuable skills and knowledge being appropriated by firms in other industries (Lorenzen & Frederiksen, 2008). Therefore, it is important to note that the effects of localization and urbanization economies could be correlated, which could lead to results where the effects of one type of externalities are dependent on the availability of the other.

This paper examines the question whether agglomeration economies affect the choice of location in establishing a new firm. In order to properly answer this question, it is vital to inspect the different approaches in documenting and quantifying the benefits of cities. It is important to mention that many of the models involving externalities and agglomeration economies describe these concepts in the context of cities, the implications are relevant for broader areas, such as regions (Fujita & Thisse, 1996).

Puga (2009) discusses the advantages of large urban environments in terms of increased productivity of workers and firms and mentions the increased innovativeness of within large cities. The author provides a good preview of the main approaches in the field of agglomeration economies. The first is to view the excessive localization of firms as a sign of agglomeration economies. The spatial impossibility theorem addresses this issue, stating that any equilibrium in the competitive environment that includes transport costs will have only small independent firms, excluding increased returns or indivisibilities (Puga, 2009; Starrett, 1978). This implies that it is impossible for large interconnected firms to cluster near each other only due to uniform distribution of firms. Duranton & Overman (2005) also look into this issue by using a distance based approach, which accounts for the entire distribution of paired distances between manufacturing plants and compares them to a distribution that is made up of random allocation of plants. Their results show that around 50% of the sectors, measured at the 4-digit level, are more localized than is possible under uniform distribution. Fujita & Thisse (1996) also address the issue that uniform distribution cannot properly produce a spatial equilibrium since any marginal diviation is enough for the population to shift towards an irregular distribuion.

The second approach is to quantify agglomeration economies though wages and rents. This method assumes that in a competitive market, workers would be paid their due marginal product. This is achieved in large and dense urban environments and as mentioned before is offset by the increased commuting and housing costs sustained by the workers. Therefore, firms have the choice to relocate should these costs are higher than the benefits of the environment. Glaeser & Maré (2001) examine the idea of increased wages

and how it may be because of the larger and more capable labour pool in cities. Their analysis shows that large urban centers exhibit a significant wage premium. Additionally, Combes, Duranton, Gobillon, & Roux (2009b) use worker fixed-effects and see a decrease in the estimates of agglomeration economies by one half (Puga, 2009).

The spatial variation of rents approach is very similar since whether firms choose to operate in a region with increased rents, there must be another advantage that counteract them. Roback (1982) separated the consumption amenities from the increased productivity in cities. Generally, productivity is assosiated with higher wages and rents, however, amenities are linked to higher rents but with lower wages since workers would be content with lower wages if there are abundant consumption amenities in the city. Therefore, the net effect can be unclear (Puga, 2009).

The third approach described in Puga (2009) is to use productivity as evidence of local increasing returns. The direct measure of agglomeration economies is to use data on outputs and inputs and estimating how productivity varies across space. However, the literature currently has moved towards using total factor productivity, calculated at the aggregate level for each area in question. Another approach in this context of measuring agglomeration economies is the used in Ciccone & Hall (1996), who instrument for size and density to account for a location's underlining productivity advantage, which would attract firms and workers and grow in a snowball effect. Overall, according to Puga (2009) the magnitudes which are presented in productivity studies are that a doubling of a city's size results in a 3-8 percent increased productivity for a substantial range of cities.

Puga (2009) also suggests a counter argument to agglomeration economies. This argument follows the idea of firm selection, where large cities with a very competitive environment incentivize firms to increase productivity otherwise they fail. Melitz & Ottaviano (2008) confirm this relationship and their results suggest that an increase in city size leads to an increase in productivity and in turn leads to less productive firms to exit the market. Combes, Duranton, Gobillon, Puga, & Roux (2012) examine both those issues and find a clear distinction in the shape of the log productivity distribution. In the case of firm selection it would be left-truncated since firms exit, while in the case of agglomeration economies, it would be right-shifted since all firms increase productivity. Overall, the authors find that productivity differenses across urban areas in France are mostly explain by agglomeration economies.

It is important to mention another approach, explained in Fujita & Thisse (1996), who explain the separation of firm activities into front office and back office. The former conducts activities such as communication with other firms while the latter does all other activities. The results show that when commuting costs decrease, the separation of business and residence areas increase. Furthermore, when intrafirm communication costs fall, the two types of offices separate, leading to front offices being

established in city centers, while back offices are in the outskirts. This separation provides an entirely different way of viewing city structure, which in turn affects the way agglomeration economies affect urban environments (Fujita & Thisse, 1996).

In relation to the literature referenced above, I look into the localization and urbanization issues in the same model and attempt to provide a relation between them and the new ventures. Furthermore, in this paper I follow the approach where excessive localization of firms is viewed as evidence of agglomeration economies as described in Puga (2009) and Fujita & Thisse (1996). The main reason for this is the ability of this approach to accommodate a large sample of firms and regions. Acquiring data on firm level for wagas and for total productivity, as well as for regional economic factors is beyond the scope of this paper. Futhermore, this approach is also most suited to answer the question: where to establish a new firm. Most research shows a positive effect of agglomeration economies on firms, however, both aspects have some possible diseconomies affecting them (increases in wages, rents, commuting, pollution).

Bosma, van Stel, & Suddle (2008) address the issue of desinction benween new subsidiaries and startups in the context of localization and urbanization in both service and manufacturing industries. Their results indicate that the urbanizaton externalities are particularly beneficial for new subsidieries, while localization effects are advantegous mainly towards independent ventures. Mukkala (2004) also find that these two types of agglomeration economies have a positive effects on firms in the Finnish manufacturing industry. Localization externalities in partucular have a very strong influence on small firms, while urbanization externalities have a more mixed effect. However, as stated by the author, the finnish urbanization is quite low and therefore it is more likely that these externalities to be more pronounced in the Netherlands. Viladecans-Marsal (2004) examines the Spannish manufacturing industry in terms of localization and urbanization and finds that both are applicable. In particular, the former influences mostly smaller firms with a highly concentrated pattern of location in industries such as textiles, leather and footwear. In contrast, urbanization economies are the most important factor for more technologically advances sectors (office and computing machinery). Figueiredo, Guimaraes, & Woodward (2009) examine aggomeration economies, while focusing on localization externalities, and establishment size. Their research also points towards Marshall externalities being vital for small firms. Overall, I expect that both types of externalities to have a positive and significant effect on the location of new firms.

H1: Localization economies are expected to have a positive and significant effect on the location of new firms.

H2: Urbanization economies have a positive and significant effect on the site of new ventures.

3. Data and Methodology

3.1. Data

The analysis in this paper mainly relies on the data from the LISA database on Dutch companies and from the Eurostat database in the period 2011-2014. The time range chosen for this paper is because the method used for acquiring the dependent variable requires several years in order to have a larger sample of observations. Furthermore, it was necessary to move away as much as possible from the financial crisis the recovery period that followed because it would not perfectly answer the research question. In the time immediately after the crisis firms and entrepreneurs would be too affected by these economic conditions when deciding where to establish a new venture to enable this analysis to accurately grasp the effects of agglomeration economies. Additionally, after 2015 the Eurostat data becomes largely unreliable for some of the control variables, such as road transport, because the database was not fully completed at the time of writing this paper.

Overall the sample used has around 896 thousand observations. However, because the aim of the paper is specifically on the four biggest regions in the Netherlands (Greater Amsterdam, Greater Rotterdam, Utrecht and The Hague), the localization measure Is computed only for them. Therefore, the final number of observations used in the analysis is 270 thousand.

As mentioned in Stavropoulos & Skuras (2016) there is no perfect scale on which to analyse agglomeration economies. In the paper, the authors use Nuts 2 level, however, they examine several countries. In the case of the Netherlands, Nuts 2 would not be ideal to examine the effects of both localization and urbanization because of the much smaller scale needed for this analysis. Therefore, Nuts 3 level is more applicable because it consists of functional regions that are much more suitable for a single country (Bosma, van Stel, & Suddle, 2008).

The industries under consideration in this paper are manufacturing and service industries. Overall, the choice of industries aimed to capture as much of the economic activity as possible, without including outliers that do not skew the results. Therefore, sectors such as mining, quarrying and construction are not included in the localization proxy. However, the agriculture sector, which is not part of the services or the manufacturing industries, is in taken into consideration because it is one of the 9 sectors that are subsidized by the Dutch government. However, the industries will be examined in more detail in the robustness checks section.

Dependent variable: The main variable of interest is a dummy showing whether a company is a new venture or not. The data is from the LISA database, which provides each independent company with its own unique number. This number is used in this paper by merging the data for two years and afterwards all companies who are only in the latter year are obviously newly created firms, while companies who are only in the former year are firms that are out of business. Furthermore, companies who continue to survive after merging the data for all four years are a proxy for successful firms. However, because of multicollinearity issues I was unable to accommodate both new firms and successful firms. The regressions with only the latter are examined in the robustness checks section.

Independent variable: The explanatory variables used in this analysis are the proxies for localization and urbanization. The former is explained using the method in (Stavropoulos & Skuras, 2016), where the authors calculated the location quotient for a firm's industry and region, using:

$$LQ_{mj} = \frac{E_{mj}}{\sum_m E_{mj}} \Big/ \frac{\sum_j E_{mj}}{\sum_j \sum_m E_{mj}}$$

Where Emj is the employment as stated in the LISA database for region j and industry m. ∑m Emj is the sum of the employment in all industries for region j, ∑j Emj is the employment for industry m in all regions, ∑j ∑m Emj is the employment in all industries and all regions (Stavropoulos & Skuras, 2016). The main advantages of using the employment location quotient are its computational simplicity and the availability of regional data by industry. However, the main disadvantage is in regards the Marshall externalities theory, since it does not distinguish between internal and external economies of scale (Figueiredo, Guimaraes, & Woodward, 2009). However, in this paper I look into only new firms, most viewed only in the year of their establishment. Therefore, the effect of internal economies of scale is hardly applicable since companies rarely have the opportunity to apply them in such a short time period.

Following (Bosma, van Stel, & Suddle, 2008), I use population density as a measure of urbanization. The variable represents the ratio between the annual average population and the land area. The land area concept (excluding inland waters) should be used wherever available; if not available then the total area, including inland waters (area of lakes and rivers) is used. This measure is taken from Eurostat at the Nuts 3 level. Kie (1997) in his paper on US manufacturing industries on a state level proposes that there are states with low population density but with large urban areas. However, this issue is not applicable to the analysis

in this paper. In this case, the urbanization density much more accurately portrays the proportion of urban areas to the overall region. The main reason is the separation into much smaller regions compared to the research in Kie (1997).

Control variables:

The control variables used in the analysis are categorized into three sections. First are the macroeconomic factors, including GDP in millions of Euro and the total population change, both taken from Eurostat. These factors aim to capture the variation stemming from the differences between the four regions under consideration. Gross Domestic Product is defined as the sum of all goods and services produced in a country over time, without double counting products used in other output. It includes the production of consumer goods and services, even government services, and investment goods. Overall, GDP on a regional level is a key variable enabling the regression to capture the economic growth in the period in question, and how this may affect the choice for a location of a new venture. The variable total population change includes both the natural population change and the net migration plus statistical adjustment. From the figure in Appendix B, a clear pattern is forming, where the four regions with highest increase in the population are Greater Amsterdam, The Hague, Utrecht and Greater Rotterdam Area. This brings an interesting point, since these regions are obviously with the largest population and with an ever-increasing population, should they be the optimal location for a new venture?

Second is the microeconomic factor, which is national annual road freight transport by regions of loading from the Eurostat database. This variable shows the total volume of goods carried by the road transport in the place of loading, measured in tonnes. This data captures: national transport; international transport; cross-trade; cabotage. Compared to previous years, cabotage is the only type of transport to have grown in 2011, while the others have severely declined (De Angelis & Roubanis, 2012). Furthermore, from the total number of relevant goods vehicles in Appendix A we see that overall the trend of overall decline continues in the following years until 2014. Heston & Lipsey (2007) explain that congestion is a main reason for the decrease of the quality of road freight transport, due to lower average vehicle speed, more traffic jams and accidents. Therefore, this variable can be viewed as another way of analyzing the urban conditions in a city.

Lastly, the Intellectual Property Rights category of controls includes the number of Community Design applications and their total number, as well as number of trademark applications and registrations. Ideally, patent applications would have been a good addition to this list, however, there were no data for 2013 and 2014 available and therefore it was excluded from this analysis. These variables attempt to explain

whether IPR affect the clustering of new ventures. The data is from the Eurostat database and is based on raw figures received from the European Union Intellectual Property Office, the EU agency responsible for registering European Union trademarks and designs. Trademarks and designs reflect non-technological innovation in every sector of economic life, including services. In this paper, these indicators are used to provide a link between non-technological innovation and the market (Eurostat, 2017).

It is important to note that from Appendix B, it is clear that the Greater Rotterdam Area has a substantial variation in the total population change indicators. In order to account for it, included in the analysis are year fixed-effects. Their aim is to capture the time differences in the model and account for changes during that period. Furthermore, Appendix C shows the Descriptive Statistics and Appendix D the variables table for greater detail on the indicators included.

3.2. Methodology

As stated before, the dependent variable is binary, therefore, in order to produce efficient estimates, I use a logit model in three configurations. The first is a basic regression with only the main variables of interest – the location quotient and population density, and the year fixed-effects. The aim is to view the relationship between the variables in its most simple form and use it as a basis for comparison with the other models. The second regression introduces GDP and population change to the analysis. This model attempts to capture the major economic and demographic conditions. Ideally, variables such as total population and geographic area would have been a viable inclusion, however, because of the nature of the urbanization measure including those variables introduces severe multicollinearity. The third model has variables for annual road transport and the IPR variables. This regression tests more specific issues i.e. congestion and the non-technological proprietary information. The goal is to examine their effect in the context of choosing a location of a new venture.

The reason for using three models is the incompatibility of the control variables. When regressed in the same model, the regression exhibits issues such as non-concavity. This stems from a singularity or near-singularity in the regression, meaning that the model is not correctly specified. Furthermore, even after separating the model, problems such as multicollinearity remained when using region specific fixed effects. In order to fix this issue, I excluded them from the model and the final models only account for the clustering in the regions. However, from VIF tests, there still remains the issue with multicollinearity in the third model and the results from it need to be taken with some skepticism.

The regression equation for all models is as follows:

$Pr(y=1|X) = \exp(\beta_0 + \beta_j X)/1+\exp(\beta_0 + \beta_j X)$

Where y is the dependent variable (new firm), X is a vector for all the explanatory variables (LQ, population density, GDP, population change, annual road loading, community design applications, trademark applications) and β_j is a vector of estimable parameters.

4. Results

As mentioned above, the analysis in this paper tests the probability of establishing a new venture against the proxies of two of the main types of agglomeration economies- localization and urbanization. The models shown in Table 1 present mixed results in regards of the effects of these factors.

In the first regression, the LQ measure is positive and significant at the 1% level, which is consistent with the first hypothesis and with the findings of the existing literature. However, the urbanization proxy is neither positive nor significant. From Table 1 it is clear that the effects are relatively small, although negative. These results propose interesting points. First, this could be because of the specific regions being analyzed, meaning that if other Dutch regions were considered, the effects could be different. Another point is that the choice of location may not be conditional on the cities themselves but on other factors and therefore the population density variable would not be able to capture the effect. This will be discussed in more detail in the Discussion Section.

The results for the localization and urbanization variables are very similar in the second regression. LQ is significant at the 1% level and positive, while population density is insignificant and negative. However, a slight increase in the p value is visible, meaning that by removing some of the variation, it becomes slightly more relevant.

The GDP variable produces an interesting result- it is significant at the 5% level and negative. This goes against the initial expectations that a region with high GDP and therefore high economic development, would increase the probability of a new venture being located in the area. However, these results coincide with the estimates of the urbanization proxy, pointing to the conclusion that it is highly possible for the biggest regions in the Netherlands to be undesirable for establishing a new firm because of the

diseconomies that are present. This is not the case for the population change variable. In the model, the coefficients are positive and significant at the 1% level, implying that an increase in the variation of a population could increase the probability of new venture opening in the region. Therefore, the economic situation seems to have less impact compared to the demographic conditions, only considering the signs of the variables.

As mentioned above, the third model has some multicollinearity issues, which are taken into account when presenting the results. In Appendix C are presented the results of a VIF test on these variables, where it is apparent that the road transport variable is most affected by multicollinearity, while density, community design and trademark application variables are only slightly influenced. The LQ measure is unaffected. Overall, the localization proxy continues to be positive and significant as in the previous models – confirming the first hypothesis entirely. The urbanization proxy also does not change sign and is negative, however, the coefficient is significant at the 1% level. This may be attributed to the multicollinearity issue. The second hypothesis is not confirmed in any of the models.

The variable of road transport is negative and significant at the 1% level, which coincides with the notion that factors which are closely related to the conditions of a city do not seem to foster the establishment of new firms. The point made by Heston & Lipsey (2007) that the decline of road transport is indicative of congestion and the overall inefficiency of this type of trasport of goods. If the road transport variable is viewed as a proxy for traffic, the result seems quite natural since it implies a high commuting costs for workers and high transportation costs for goods. However, the variable could also be indicative of the lower value that road freigh trasport has compared to train, water or air transport for businesses, which does not nessecarily affect the location of new companies.

The IPR variables are both positive and significant at the 1% level. However, their coefficients are close to zero, which means that the relationship between them and the location of a new firm does not have much weight to it. Overall, it seems that the number of non-technological intelectual property applications is not relevant for new ventures. This proposes an interesting implication- even in businesses where these types of proprietary information is vital, new firms could have a different approach towards drawing benefits from them. Additionally, it is important to note that this type of IPR is rarely central to a firm's value proposition, meaning that it is more likely for these factors to not be considered when choosing a location of a new venture.

The year fixed-effects in all three models show a general trend of the later years (2013,2014) to be positive and significan relative to 2011, which is the reference category in the regressions. This can be attributed to the growing number of firms and be indicative of good economic conditions in the Netherlands. Additionally, these results could be indicative of a recovery period after the financial crisis.

Table 1

VARIABLES	(1) Basic Model	(2) Macro Variables	(3) Road Transport and IPR
LQ	1.320***	1.148***	1.198***
	(0.063)	(0.181)	(0.167)
Urbanization	-0.027	-0.116	-1.017***
	(0.141)	(0.080)	(0.055)
GDP		-0.264**	
		(0.106)	
Population change		0.447***	
		(0.144)	
Road transport			-0.934***
			(0.052)
Community Design			0.004***
			(0.001)
Trademark			0.001***
			(0.000)
2012	-0.045	0.077	-0.130**
	(0.100)	(0.118)	(0.051)
2013	0.274**	0.358***	0.322***
	(0.117)	(0.133)	(0.090)
2014	0.544***	0.515***	0.490***
	(0.152)	(0.176)	(0.100)
Constant	-0.735	-1.116	15.229***
	(1.100)	(1.346)	(0.890)
Observations	269,380	269,380	260,094

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

5. Robustness Checks

In this section are included all the robustness check I did on the analysis. Overall, there are two general aims. First, to test the results of new ventures compared to the results of the regression with a proxy for successful companies. Second type of checks concern the distinction between the industries in the Netherlands.

The proxy for successful firms in the context of this analysis is firms who have survived their first year. In Appendix D is included a frequencies table for the two variables. The number of successful firms is much smaller, as to be expected. In Appendix E are included the same logit regressions as in the main analysis, but with the proxy for successful firms as dependent variable. Overall, the results are largely the same. The location quotient is positive and significant at 1% level in all three regressions. The GDP, population change, road transport and IPR variables do not have any noteworthy changes to their sign or significance. The most notable difference is the urbanization proxy (population density). In the first and second models, the coefficients are now positive, as well as significant in the second regression. This greatly contrasts the results of the regressions with new firms. One possible explanation is that only firms who can survive their first year are able to benefit from urbanization economies. This will be elaborated in greater detail in the discussion section.

The industries related robustness checks distinguish between manufacturing and service industries. The manufacturing sector is separated into high-tech, medium-low tech and low-tech according to OECD classification by knowledge spillovers (OECD , 2011). The high-tech sector includes both the high and medium-high tech, because by themselves the regression had too few observations. The services industries are divided into the 9 sectors subsidized by the Dutch government. These sectors are the Agriculture, Horticulture, High Tech, Energy, Logistics, Creative Industry, Life Sciences, Chemicals and Water (van den Berge & Snoei, 2017). However, the horticulture sector was not included because it is not in the SBI industries index. The regressions are run with the same coefficients as the basic model in the results section.

Overall, 13 regressions were run: manufacturing and the three subsectors, services and its 7 subsectors, as well as agriculture. However, the regressions for chemistry, energy and water industries had too few observations and therefore were not worth considering. As presented in Appendix E, the results are very mixed. The manufacturing industries models showed no clear pattern. In the all manufacturing, high-tech

and low-tech models both measures for agglomeration economies are insignificant. Only in the medium-low manufacturing sector, the LQ and density variables are significant but negative. This suggests that in this sector, these types of agglomeration economies do not help new ventures, meaning that potential entrants would steer clear of these areas.

The services industries account for a much larger percentage of all observations and naturally have results more in line with the main finding. LQ is positive and significant at 1% level in all sectors. The urbanization proxy is insignificant and negative in Creative Industries and Life Sciences, negative and significant in the High tech service sector and positive and insignificant in the Logistics and all service regressions. The agriculture regression presents a bit different results – both are insignificant but LQ with negative sign and density with a positive sign, which is the same as all manufacturing model.

Overall, no clear conclusion can be drawn on these results, but there are some patterns. Agglomeration economies seem to be largely unimportant when considering the manufacturing sectors, however, it is important to note that the results may be skewed because of the small amount of observations. In contrast, the service industries the localization proxy is very relevant and the urbanization is only in some of the models. There results tend to coincide with the main findings of this paper in the sense that localization is proven to be heavily supported by the regressions, while urbanization is found to be insignificant and negative in most.

6. Discussion

As stated before, my first hypothesis was confirmed by both my research and the existing literature. However, Bosma, van Stel, & Suddle (2008) present interesting finding with their analysis of agglomeration economies on new firms, which are separated into new subsidiaries and independent companies. The authors use a similar setting for their research to the one used in this paper. Both examine the regions in the Netherlands in terms of localization and urbanization economies on a Nuts 3 level and in regards to manufacturing and services industries. Consequently, the result that Marshall externalities have insignificant effect on new subsidiaries for both manufacturing and services sectors is surprising. However, localization externalities are present for independent firms, therefore, it is possible that the data used in

this analysis has a much larger percentage independent companies and the significance of the LQ represents this relationship. The same argument could be made when comparing Bosma, van Stel, & Suddle (2008) results on urbanization, which is found to be mostly beneficial for subsidiaries. Therefore, most of the implications made for independent firms by their paper coincide with the results from my analysis.

Much of the research shows that urbanization economies should be beneficial for firms, there are some instances where that is not the case. Henderson (2003) examines both Marshall and Jacobs externalities and the results confirm my results. Localization economies have a positive and significant effect on the industry, while urbanization does not. Moreover, even distinguishing between static and dynamic externalities does not change this result. Another possible reason for the negative urbanization ceofficients is the diseconomies of scale present in any big cities. As mentioned above, variables such as GDP and road transportat, which accounts for the economic situation, have negative effect on the choice of a location for new firms. this contrasts heavily from the result of population change. Bosma, van Stel, & Suddle (2008) draw their independent firms sample from the labour pool and produce positive and significant results, therefore, it seems natural to assume that an increase in population would have a positive effect.

Furthermore, even though Mukkala (2004) presents positive effects of urbanization in some sectors, in the paper these externalities are found to be important mainly for large firms. This could explain the negative and insignificant effects of the density measure in this analysis since only very young firms are considered. Furthermore, this could be viewed as an explanation for the results of the regressions with the proxy for successful firms. Generally, firms who survive their first year are much larger in terms of both workers and revenue. Therefore, it seems natural for the bigger firms examined in the robbustness checks to be more likely to benefit from urbanization. Another way to view this is to expect a treshold for new ventures which needs to be surpassed in order to benefit from urbanization economies. Consequently, only the successful firms in the sample would be able to surpass it. Therefore, it could be the case that it is not the choice of where to establish a company but the ability to grow enough to benefit.

Another key point is the distinction between services and manufacturing industries. Overall my results show that both urbanization and localization could have a positive effect but in manufacturing models this is largely insignificant, while in services models only the urbanization measure is insignificant. Bosma, van Stel, & Suddle (2008) find that manufacturing sector should benefit more from agglomeration economies compared to services, due to the knowledge intensity. This is not completely rejected by my analysis, since no clear comparison can be made in this paper due to the logit models used. Furthermore, noteworthy is to mention that their research is on the period 1980 -2002, which covers admittedly a quite larger range of

but is much earlier in time. Therefore, the differences in my finding and the ones by Bosma, van Stel, & Suddle (2008) could be attributed in no small part to the progress of time. Another point is the Dutch government's subsidies, which are mostly aimed at the services sectors. Introduced in 2011, this policy could greatly influence the results because in the first few years since its implimentation, many entrepreneurs in the Netherlands would want to be the first to benefit from it, leading to a significant bias in the results.

7. Conclusion

In conclusion, using data on the four biggest regions in the Netherlands on Nuts 3 level, the paper attempts to examine the choice of a new location for a firm regarding localization and urbanization economies. Further analysis is done to distinguish between successful firms and across industries (manufacturing and services). The main results fully support the notion that localization economies are extremely important in choosing a location. In contrast, urbanization economies are found to be largely insignificant and even negatively related to the new firms. The robustness checks show that successful firms also benefit from Marshall externalities but also are more likely to take advantage of Jacobs externalities as well. However, no clear pattern has been established when considering the manufacturing, services industries and their subsectors, only the tendency of localization economies to be positive and sig in the services sectors.

This paper also has its limitations, the main of which is the type of urbanization measure used – population density. In this analysis, the biggest issue with density is the correlation with the other variables, resulting in multicollinearity problems. This in turn leads to another major limitation of this paper, which is the lack of control variables. Many of the region-specific indicators are highly linked to density and would increase the multicollinearity problem. Additionally, there is the issue with adding variables such as presence of universities, infrastructure and local institutions – the possible reverse causality. These factors prove a big challenge and more research is required in order to causally be able to link agglomeration economies and productivity (Groot, Groot, & Smit, 2014).

Another possible limitation is the usage of area-based approach. According to papers such as Duranton & Overman (2005) and Huisman & van Wissen (2004), the better approach to agglomeration economies is a distance-based one. However, it is exceedingly difficult to examine both types of agglomeration economies using a distance-based approach, as well as account for the high specificity in the data required. Therefore, this issue should be a major point of further research because of the promising nature of the method.

Furthermore, there has been some research using infrastructure variables in examining firm productivity. However, a small part of the existing literature has linked it to agglomeration economies and new firms (Elberts & McMillen, 1999). By ommiting infrastructure indicators from the model, studies assume there is no significant difference in services from capital stock across cities. Therefore, if infrastructure per capita is increased in large cities or in expecially congested ones, when comparing to smaller or less congested cities the estimates could be positively biased (Elberts & McMillen, 1999). The results in this paper also suggest that an important connection exists between agglomeration economies and infrastructure, which require further research to be done.

Overall, this paper brings to the existing literature a further argument of the significant effects of localization economies, exhibiting their effect on newly established firms. Additionally, it sheds light to the uncertain issue of urbanization economines, especially when using population density as a measure, and the research needed on the topic.

References

Bosma, N., van Stel, A., & Suddle, K. (2008). The Geography of New Firm Formation: Evidence from Independent Start-ups and New Subsidiaries in the Netherlands. *International Entrepreneurship and Management Journal*, 129-146.

Ciccone, A., & Hall, R. (1996). Productivity and the density of economic activity. *American Economic Review*, 54–70.

Combes, P., Duranton, G., Gobillon, L., & Roux, S. (2009). Estimating agglomeration effects with history, geology, and worker fixed-effects. *The Economics of Agglomeration*.

Combes, P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 2543-2594.

De Angelis, L., & Roubanis, N. (2012, 09 21). *Decline in European road freight transport in 2011 reflecting the economic climate.* Retrieved from Eurostat website: http://ec.europa.eu/eurostat/en/web/products-statistics-in-focus/-/KS-SF-12-038

Duranton, G., & Overman, H. (2005). Testing for localization using microgeographic data. *Review of Economic Studies*, 1077-1106.

Elberts, R., & McMillen, D. (1999). Agglomeration Economies and Urban Public Infrastructure. *Handbook of Regional and Urban Economics*, 1455-1495.

Eurostat. (2017, 03 06). *Intellectual property rights statistics*. Retrieved from Eurostat Statistics Explained: http://ec.europa.eu/eurostat/statistics-explained/index.php/Intellectual_property_rights_statistics

Figueiredo, O., Guimaraes, P., & Woodward, D. (2009). Localization economies and establishment size: was Marshall right after all? *Journal of Economic Geography*, 853–868.

Fujita, M., & Thisse, J. (1996). Economics of Agglomeration. *Journal of Japanese and International Economies*, 339-378.

Glaeser, E., & Maré, D. (2001). Cities and skills. *Journal of Labor Economics*, 316-342.

Groot, S., Groot, H., & Smit, M. (2014). Regional wage differences in the Netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, 503-523.

Henderson, J. (2003). Marshall's scale economies. *Journal of Urban Economics*, 1-28.

Heston, A., & Lipsey, R. (2007). *International and Interarea Comparisons of Income, Output, and Prices.* Chicago : University of Chicago Press.

Holm, U., & Pedersen, T. (2000). *Centres of Excellence.* London: Macmillan.

Huisman, C., & van Wissen, L. (2004). Localization effects of firm startups and closures in the Netherlands. *The Annals of Regional Science*, 291-310.

Jacobs, J. (1969). *The Economy of Cities,.* Random House.

Kie, C. (1997). *The effects of flexible manufacturing contexts: Geographical patterns in the labor productivity impact of output growth.* Taylor & Francis.

Lorenzen, M., & Frederiksen, L. (2008). Why do Cultural Industries Cluster? Localization, Urbanization, Products and Projects. *Creative Cities, Cultural Clusters, and Local Economic Development*, 155-179.

Marshall, A. (1890). *Principles of Economics.* London: Macmillan.

Melitz, M., & Ottaviano, G. (2008). Market size, trade and productivity. *Review of Economic Studies*, 295–316.

Mukkala, K. (2004). Agglomeration economies in the Finnish manufacturing sector. *Applied Economics*, 2419-2427.

OECD . (2011, 7 7). *Directorate for Science, Technology and Industry Economic Analysis and Statistics Division.* Retrieved from OECD website: https://www.oecd.org/sti/ind/48350231.pdf

Porter, M. E. (2000). Locations, clusters and company strategy. *The Oxford handbook of economic geography*, 253-274.

Puga, D. (2009). The magnitude and causes of agglomeration economies. *Journal of Regional Science*.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 1257-1278.

Starrett, D. (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 21-37.

Stavropoulos, S., & Skuras, D. (2016). Firm Profitability and Agglomeration Economies: an Elusive Relationship. *Tijdschrift voor economische en sociale geografie*, 66-80.

van den Berge, H., & Snoei, W. (2017). *Doing business in The Netherlands Subsidies and Financing.* Visser&Visser.

Viladecans-Marsal, E. (2004). Agglomeration Economies and Idustrial Location: City Level Evidence. *Journal of Economic Geography*, 565-582.

Appendices

Appendix A

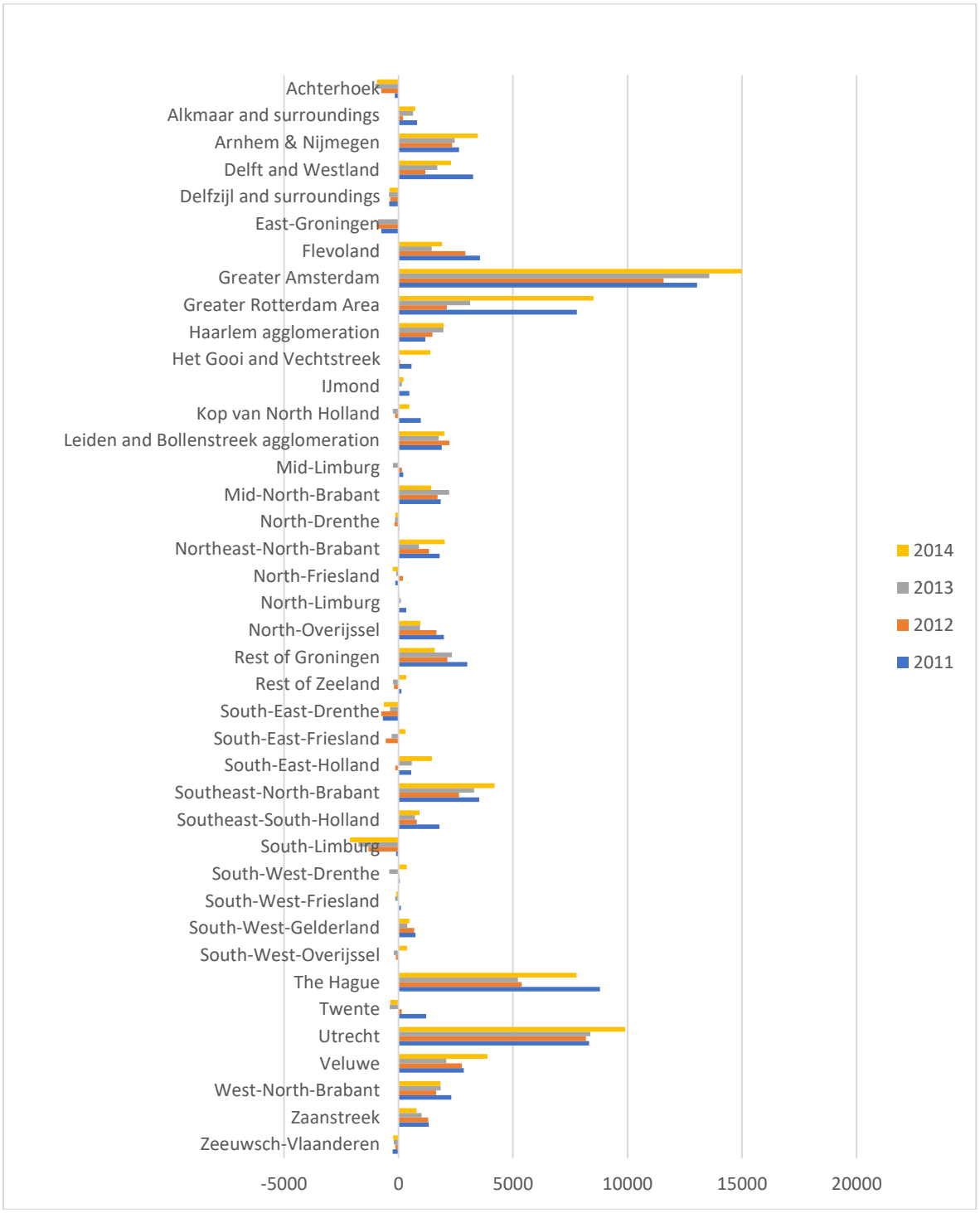
Total number of relevant goods vehicles in there porting countries

	2011	2012	2013	2014	2015	Change 2014-2015 (%)
EU-28 (1)	4 350 366	4 212 055	4 155 046	4 132 242	4 128 310	-0.1
Belgium	102 843	125 744	97 936	114 907	105 930	-7.8
Bulgaria	137 558	137 987	138 213	128 066	129 455	1.1
Czech Republic	122 608	118 166	124 872	132 355	140 639	6.3
Denmark	39 858	39 044	46 785	36 996	36 813	-0.5
Germany (2)	478 535	483 021	487 036	500 275	500 275	0.0
Estonia	17 096	17 439	17 698	17 963	16 345	-9.0
Ireland	80 483	78 847	75 692	82 247	85 952	4.5
Greece	121 179	121 301	122 772	112 125	115 403	2.9
Spain	354 723	337 165	318 167	306 724	308 498	0.6
France	545 759	535 908	542 784	541 312	531 932	-1.7
Croatia	27 193	25 829	25 203	24 654	25 729	4.4
Italy (2)	293 536	282 174	255 605	229 790	229 790	0.0
Cyprus	13 431	13 325	13 068	12 781	12 466	-2.5
Latvia	18 526	19 740	20 360	20 573	20 192	-1.9
Lithuania	38 001	38 331	39 506	40 273	37 078	-7.9
Luxembourg	10 702	9 721	9 694	9 528	9 511	-0.2
Hungary	75 141	73 773	73 288	74 876	76 488	2.2
Malta (4)	:	:	:	:	:	:
Netherlands	286 816	141 858	133 337	130 620	98 484	-24.6
Austria	68 490	67 877	66 749	66 103	65 750	-0.5
Poland	626 289	645 285	649 985	649 553	666 153	2.6
Portugal	110 971	113 597	108 287	99 152	98 783	-0.4
Romania	74 529	85 125	88 586	92 758	97 506	5.1
Slovenia	23 190	22 242	21 814	22 046	22 821	3.5
Slovakia	126 152	129 951	132 908	134 220	137 617	2.5
Finland (3)	98 719	101 679	102 774	104 083	104 997	0.9
Sweden	61 260	61 731	61 002	61 115	61 521	0.7
United Kingdom (3)	396 780	385 200	380 929	387 151	392 185	1.3
Norway	42 290	42 926	43 607	44 093	39 816	-9.7
Switzerland	49 895	50 459	50 419	50 630	51 915	2.5

(1) EU-28: provisional data for reference year 2015.
(2) DE and IT: 2014 data was used for reference year 2015.
(3) FI and UK: data includes only vehicles performing national transport
(4) Malta excluded (see chapter "data sources and availability")

Source Eurostat
http://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Total_number_of_relevant_goods_vehicles_in_the_reporting_countries,_2011-2015.png&oldid=310964

Appendix B



Bar graph presenting the population change values for the Dutch regions for each of the 4 years.

Appendix C

Descriptive statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
LQ	269,380	1.041111	.1160288	.149598	1.183973
Density	895,981	6.528988	.7976812	4.973279	8.108684
GDP	895,981	9.996107	.9353713	7.181592	11.4443
Road Transport	875,203	9.588138	.7043062	7.095064	11.08756
Population Change	751,373	7.893506	1.403533	1.098612	9.615405
CD applications	880,705	43.1193	43.92963	1	188
TM applications	895,513	216.8153	243.4489	1	819

Appendix D

Variables table

Variable	Description
Location Quotient	Calculated using employment statistics for each region and industry in question.
Density	The ratio between the annual average population and the land area of a region, in natural logarithm.
GDP	Gross domestic product (GDP) at current market prices by NUTS 3 regions, in natural logarithm
Population Change	Total population change at Nuts 3 level, in natural logarithm
Road Transport	National annual road freight transport by regions of loading (NUTS 3), in natural logarithm.
CD applications	Number of Community Design applications per region.

TM applications	Number of Trademark applications per region.

Appendix E

VIF Statistics

Variable	VIF	1/VIF
LQ	1.21	0.823615
Population density	6.87	0.145658
Road Transport	10.09	0.099063
CD applications	6.35	0.157477
TM applications	4.26	0.234758

2012	1.56	0.642203
2013	1.52	0.656343
2014	1.51	0.662604

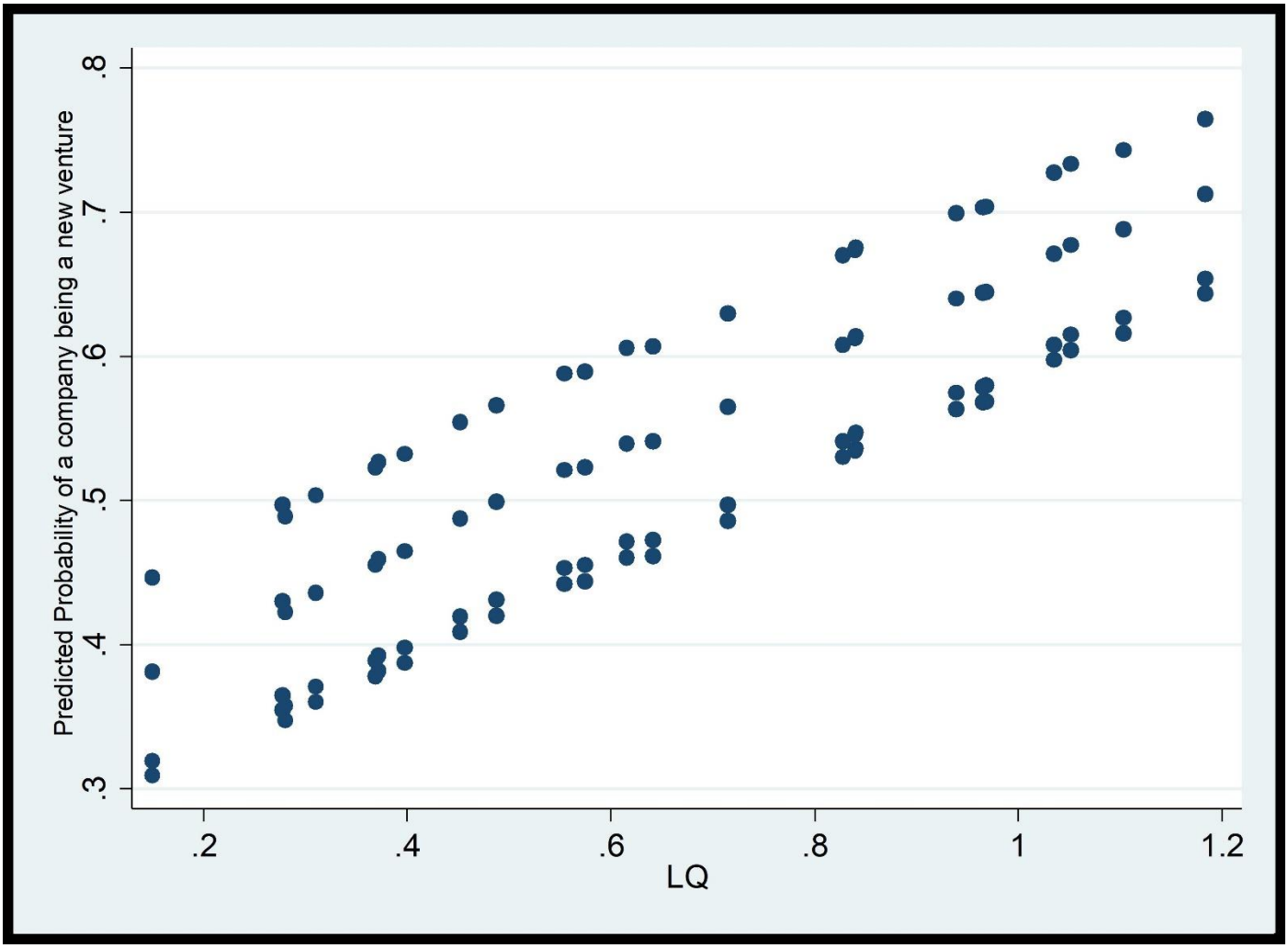
Appendix F

Frequencies tables for new and successful firms.

New firms	Frequency	Percent	Cumulative
0	331,349	36.98	36.98
1	564,632	63.02	100.00
Total	895,981	100.00	

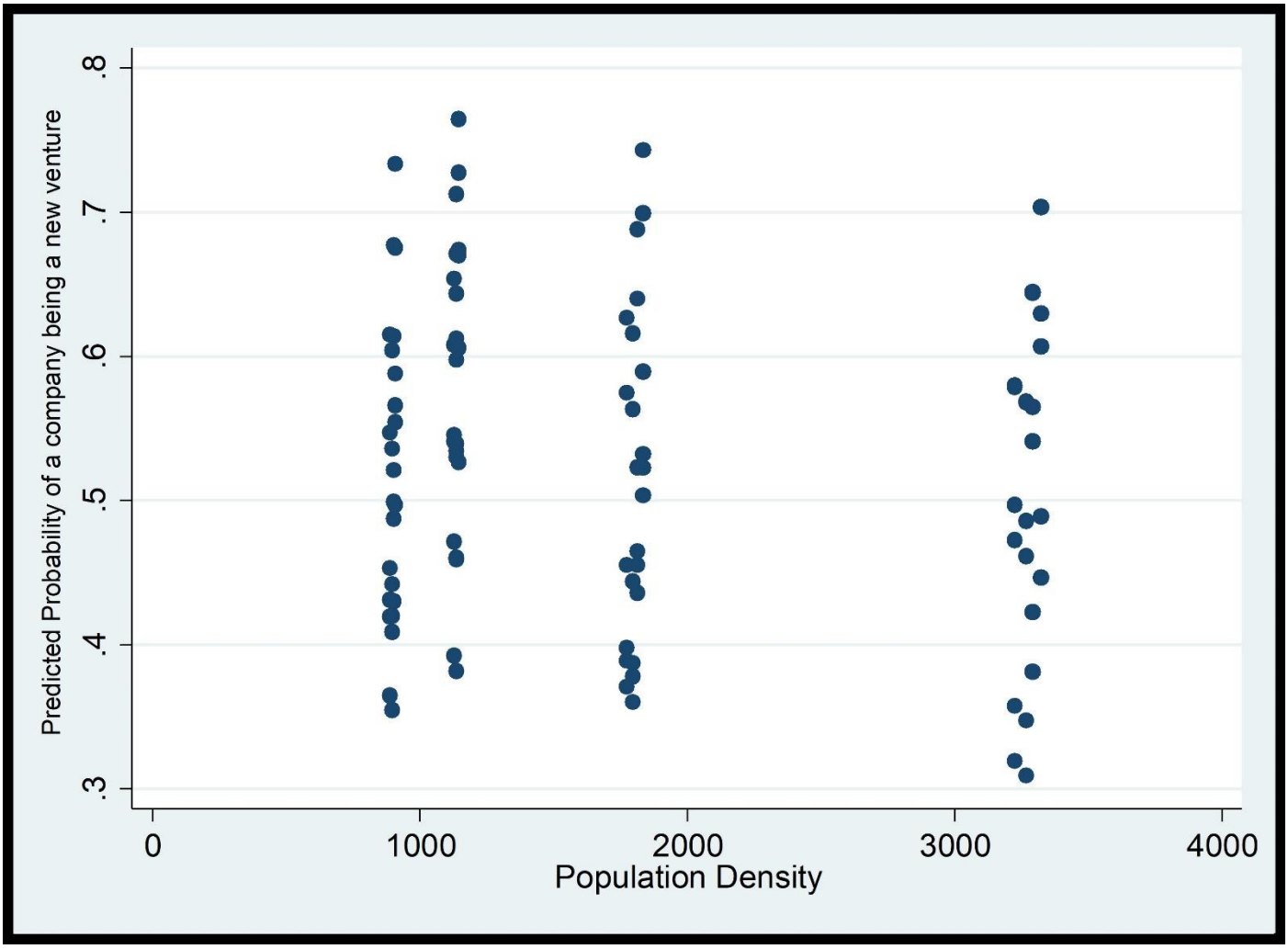
Successful firms	Frequency	Percent	Cumulative
0	783,700	87.47	87.47
1	112,281	12.53	100.00
Total	895,981	100.00	

Appendix G



This scatter plot graphs the Location quotient withthe predicted probability of a firm being a new venture. A clear pattern is formed, showing the positive relationship described in the results section.

Appendix H



This figure shows the urbanization proxy-population density against the predicted probability of a firms being a new venture.The four clusters of points represent the four regions in question. Overall,no clear positive or negative relation can be drawn, but it is evident that all the regions have different values for density but similar effects compared to the dependent variable ofthe main regresions.

Appendix J

Robustness Check- Successful firms proxy

VARIABLES	(1) Basic Model	(2) Macro Variables	(3) Road Transport and IPR
LQ	1.258***	1.192***	1.230***
	(0.204)	(0.127)	(0.111)
Urbanization	0.221	0.097*	-0.568***
	(0.139)	(0.053)	(0.051)
GDP		-0.408***	
		(0.112)	
Population change		0.482***	
		(0.167)	
Road transport			-0.746***
			(0.047)
Community Design			0.001
			(0.001)
Trademark			0.001***
			(0.000)
2012.year	-0.260***	-0.142	-0.368***
	(0.060)	(0.114)	(0.025)
2013.year	-0.879***	-0.800***	-0.889***
	(0.049)	(0.090)	(0.070)
Constant	-4.165***	-3.135***	8.680***
	(1.256)	(0.806)	(0.853)
Observations	206,414	206,414	197,128

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Appendix J
Robustness check- Industries

VARIABLES	(1) All Manufacturing	(2) High-Tech	(3) Medium-low Tech	(4) Low-Tech	(5) Services	(6) Logistics	(7) Life Sciences	(8) Creative Industry	(9) High-Tech Services	(10) Agriculture
LQ	-0.033	0.527	-0.407***	-0.883	2.404**	5.729***	3.349***	2.278***	1.699***	-0.014
	(0.438)	(0.371)	(0.035)	(0.994)	(1.123)	(1.723)	(1.131)	(0.764)	(0.180)	(2.151)
Urbanization	0.058	0.090	-0.128**	-0.185	0.002	0.484	-0.201	-0.020	-0.256***	0.368
	(0.129)	(0.288)	(0.050)	(0.211)	(0.187)	(0.325)	(0.128)	(0.123)	(0.014)	(0.829)
2012	-0.178	-0.392***	-0.261	-0.104	-0.048	0.153	-0.006	-0.110	-0.041	0.073
	(0.130)	(0.142)	(0.190)	(0.155)	(0.097)	(0.224)	(0.074)	(0.108)	(0.338)	(0.249)
2013	0.169	0.277***	-0.052	0.239	0.272**	0.620***	0.032	0.247**	0.152	0.410
	(0.174)	(0.070)	(0.292)	(0.154)	(0.110)	(0.178)	(0.106)	(0.113)	(0.167)	(0.376)
2014	0.340	0.275*	-0.187	0.533**	0.539***	0.635*	0.510***	0.570***	0.582**	0.624*
	(0.254)	(0.149)	(0.339)	(0.222)	(0.148)	(0.343)	(0.132)	(0.137)	(0.254)	(0.342)
Constant	-0.258	-0.861	1.307***	2.123	-2.100	-9.575**	-1.359	-1.530	0.482**	-3.163
	(0.907)	(2.070)	(0.302)	(2.097)	(2.492)	(4.067)	(1.848)	(1.654)	(0.240)	(5.389)
Observations	5,940	909	1,418	3,613	258,017	4,314	14,049	40,422	895	5,423

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10