The Joint Influence of Agglomeration Economies and Geographical Centrality on Startup Performance in the Netherlands

An Empirical Investigation Leveraging Crunchbase Data

By

Thijs Wentzel (S1084886)

Radboud Universiteit

Project duration:	December 2022 – June 2023	E-mail: thijs.wentzel@ru.nl	
Thesis committee:	Dr. M. de Rochemont (Maurice)	Supervisor	External
	Dr. ir. N.G. Migchels (Nanne)	Second reader	Radboud University

PREFACE

The journey of crafting this thesis has been both challenging and rewarding, characterized by in-depth learning and personal growth. Through this endeavor, I have delved deep into the fascinating world of entrepreneurship, specifically focusing on the effects of geographical factors on startup performance. This trajectory has significantly enriched my understanding of the entrepreneurial landscape and the dynamics in it.

This journey, while independently started, has been significantly shaped by the support and guidance of those around me. I would like to express my gratitude to my direct supervisor, Maurice de Rochemont, whose expertise and insights have been instrumental in guiding me throughout this process. His consistent feedback, patience, and support have contributed greatly to this end result.

I also extend my appreciation to my second reader, Nanne Migchels, whose constructive critiques and valuable perspectives have also contributed to the refinement of this thesis. Their input has encouraged me to approach my research from various angles, adding depth as well as breadth to my understanding of the subject. I hope is that the insights presented here will contribute modest yet meaningfully to the understanding of location-based factors influencing startup performance.

Thijs Wentzel,

23th of June 2023

ABSTRACT

This study examines how agglomeration economies, gauged by urban density, and geographical centrality (jointly) affect startup performance in the Netherlands, using a dataset of 727 startups and evaluating their performance on total funding and total revenue. Core theories like the Central Places Theory and Agglomeration Economies Theory are utilized. Urban density, representing agglomeration economies, significantly predicts funding but not revenue, meaning denser urban startups attract more funding without necessarily earning more revenue. Surprisingly, geographical centrality showed no significant effect on startup performance, suggesting the traditionally assumed benefits of central locations might be overestimated. However, different patterns emerge in specific industries, highlighting the complexity between location factors and industry characteristics.

Furthermore, the interaction of agglomeration economies and geographical centrality can compound the benefits of these two. Startups in less urban but central locations secured more funding, indicating an interplay between these factors. This research delivers valuable insights for startups, investors, policymakers, and academics, in providing a deeper understanding of how location-dependent factors influence startup performance. It also advocates further exploration into this relationship across various geographical contexts and industry-specific dynamics using different performance metrics to enrich the entrepreneurial literature.

Keywords: *Startups, Urban Density, Geographical Centrality, Startup Performance, Total Funding, Total Revenue, Agglomeration Economies Theory, Central Places Theory, Netherlands, Entrepreneurship, Urban Economy, Polycentric Cities, Entrepreneurial Ecosystems*

TABLE OF CONTENTS

1. INTRODUCTION

The location's impact on startup performance is a subject of ongoing debate and interest in academia and industry (Minai and Lucky, 2011; Li et al., 2016; Frenkel, 2012; Oerlemans and Meeus, 2005). Strongly urbanized areas, with their dense networks of resources and opportunities, are often considered as attractive ecosystems for startups (Minai and Lucky, 2011). This is due to their institutional environment and demographic structure (Pan and Yang, 2018; Henderson, 2002). Conversely, research has highlighted the potential for success in rural or less urbanized areas, emphasizing the value of localized knowledge, lower competition and unique market opportunities (Li et al., 2016; Schutjens and Wever, 2005).

1.1 AGGLOMERATION ECONOMIES AND STARTUP PERFORMANCE

Agglomeration economies are fundamental to the understanding of how startup performance is influenced by location. This theory refers to the benefits that firms reap from their proximity to one another, concentrating in urban areas, or 'agglomerations'. Agglomeration economies traditionally generate three types of externalities: localization, urbanization, and diversification externalities (Beaudry and Schiffauerova, 2009; Rosenthal and Strange, 2004). Localization externalities pertain to the benefits reaped within a particular industry due to the proximity of similar firms (Marshall, 1920). Diversification externalities, on the other hand, concern the innovation and cross-fertilization of ideas resulting from inter-industry interactions within an urban area (Jacobs, 1969). Finally, urbanization externalities, which is usually captured by the overall density of economic activity, relate to the advantages derived from the density of an urban area, irrespective of industry specificity (Ciccone and Hall, 1996). In this research, the primary focus will be on urbanization externalities. Specifically, this research will investigate how urban density affects startup performance, enhancing the understanding of this externality.

Such urban dense areas allow startups to access a larger and more diverse pool of resources, including human capital, knowledge, and financial resources (Schutjens and Wever, 2005; Cavallo, 2020; Tunberg, 2014). Urban environments enable closer and more efficient interactions with suppliers, service providers, and customers (Storper and Venables, 2004). Additionally, agglomerations offer key benefits such as shared public goods and infrastructure, rapid knowledge exchange, and increased innovation through spillover effects (Cavallo, 2020; Ghio et al., 2015; Jacobs 1969).

Although agglomeration economies can offer startups a wealth of opportunities and advantages, they can also present certain challenges, such as higher costs of operation, heavier competition, and congestion (Porter, 1998; Duranton and Puga, 2000; Li et al., 2016). Furthermore, recent developments in technology and infrastructure may have begun to alter the dynamics of these agglomeration economies, potentially enabling startups to thrive beyond the large metropolitan cities (Bel and Fageda, 2008). In fact, the study of Dahl and Sorensen (2012) showed that startups that were regionally embedded, rather than located in urbanized areas, had better performance outcomes.

1.2 GEOGRAPHIC CENTRALITY AND STARTUP PERFORMANCE

While substantial evidence suggests a positive impact of agglomeration economies on startup performance, the exact magnitude of this relationship remain unclear. The Central Places Theory (CPT) is another fundamental pillar that allows this research to explore the interplay between agglomeration economies and startup performance. The core principle of this theory revolves around the idea of 'geographical centrality' (Christaller, 1933). The Central Places Theory posits that the economic landscape is not uniform but is characterized by varying degrees of centrality. The spatial hierarchy significantly affects market accessibility, as more central locations have extensive reach and cater to wider markets (Eaton and Lipsey, 1982).

Such central locations, according to the theory, attract a higher degree of economic activity due to the benefits they provide. These benefits include easier access to essential resources such as labor, capital, and knowledge networks (Mulligan, 1984). Central locations offer startups similar benefits as urban areas (agglomeration economies). However, the impact of geographical centrality on startup performance is not straightforward. As Gallagher and Miller (1991) found significantly more fast-growing companies in central areas of the UK, while Stam (2005) and Anderson et al. (2005) found no statistically significant performance differences among firms in central, accessible-central, and remote-periphery locations.

1.3 RESEARCH GAP AND QUESTIONS

The current literature diverges on the factors contributing to startup performance. Some studies point towards the benefits of the agglomeration perspective, emphasizing the advantage for startups located in areas with rich resources (Mulligan, 1984). In contrast, others support the embeddedness perspective, arguing that startups thrive when they can leverage local resources, contributing to their performance (Kalantaridis, 2009). Despite these observations, the specific impact of agglomeration economies on startup performance remains unclear, with a lack of clarity surrounding the magnitude and significance of this effect. This observation is echoed by Guzman (2019), who emphasized the need for more research to evaluate the impact of urban agglomerations on startup performance.

This research goes beyond examining the direct relationship and investigates how agglomeration economies and geographic centrality jointly influence startup performance, including differences between urban dense areas in the 'center' and the 'periphery'. This multi-faceted approach aims to provide a more comprehensive understanding of the factors influencing startup success. Therefore, the research addresses the following research question:

- *"What are the effects of agglomeration economies and geographic centrality on startup performance, and does the geographic centrality influence the relationship between agglomeration economies and startup performance?"*

1.4 CONTRIBUTIONS AND LIMITATIONS

This research extends the current body of literature by exploring the relationship between agglomeration economies, geographical centrality, and startup performance. This analysis conducted in the unique context of the Netherlands offers novel insights into the influence of location-dependent factors on startup performance in a densely populated and compact country.

This study's methodological strength lies in the utilization of a comprehensive dataset of 727 Dutch startups. This large dataset enhances the potential generalizability of the study's conclusions (McEwan, 2020). The findings of this research have significant implications for entrepreneurs in the Netherlands, providing insights for aspiring entrepreneurs determining the optimal startup location and for investors and policymakers seeking a deeper understanding of the role of agglomerations in driving startup performance.

Despite its contributions, this study acknowledges limitations such as potential multicollinearity issues, the use of total funding and total revenue as performance indicators, and reliance on estimated revenue values. Additionally, this research primarily focuses on the role of agglomeration economies (measured by urban density) and geographical centrality, potentially downplaying the importance of other significant factors that influence startup performance, such as incubation participation, access to venture capital and other organizational or environmental factors (Qian et al., 2011; Jeong et al., 2019; Hansen and Wernerfelt, 1989).

2. EXISTING LITERATURE

This chapter reviews the literature on the most important concepts of this study: startup performance, agglomeration economies and geographic centrality. The literature review focuses on explaining why these concepts together shape startup performance. After reading this chapter, the reader will have a clear understanding of the current knowledge on the topic, providing a foundation for the development of the conceptual model in the subsequent chapter.

2.1 STARTUPS AND FIRM PERFORMANCE

In the realm of business, startup performance stands out as distinct from traditional small and medium enterprises (SMEs) due to its unique attributes. Startups are characterized by their recent establishment, emphasis on innovation, exposure to risk and uncertainty, and potential for rapid growth (Ehsan, 2021).

Researchers use distinct measures to assess startup performance, different from those used to evaluate firm performance in general. While firm performance in business and management research commonly focuses on indicators such as efficiency, growth, and profitability (Murphy et al., 1996), the evaluation of startup performance include startup funding, employment growth, market share, gross profit, cost control, revenue, survival rates, and successful exits (van Rijnsoever et al., 2016; Eveleens et al., 2017). Quantitative measures provide in general valuable insights, but qualitative indicators are also important and offer a deeper understanding of customer behavior, enthusiasm, and relationship dynamics between startups and their customers. By both considering qualitative and quantitative aspects startup performance assessment becomes more holistic (Jarvis et al., 2000). For measuring startup performance, additional challenges emerge. Startups are often privately held and thus not required to report their performance to public shareholders, making it difficult to obtain data (Eveleens et al., 2017).

According to Schutjens and Wever (2005), the determinants of startup performance include human capital factors (such as entrepreneurial experience and level of education), firm-associated factors (such as startup capital, ownership structure, and financial basis), and external factors (such as market concentration, dynamism, and growth). These factors collectively shape the success or failure of startups. Schutjens and Wever (2005) further argue that location emerges as a critical determinant of startup performance. Favorable regional factors, such as resource availability, accessibility, and agglomeration effects, contribute to startup success (Reynolds et al., 1994; Storey, 1994).

2.2 AGGLOMERATION ECONOMIES

Agglomerations, as described by Duranton and Puga (2004), refer to the clustering of people/firms in cities despite the availability of open space. “Agglomeration externalities,” as highlighted by Neffke et al. (2011), are crucial in understanding regional development and firm performance, as it refers to the benefits and costs that businesses gain from locating close other economic actors (Goodall, 2013). These benefits are generally classified into three types: localization, diversification, and urbanization externalities (Beaudry and Schiffauerova, 2009; Rosenthal and Strange, 2004).

2.2.1 LOCALIZATION EXTERNALITIES

Localization externalities arise from the geographic concentration of companies in the same industry, resulting in static and dynamic benefits (Duranton and Puga, 2004). Porter's clustering theory complements this concept, emphasizing the role of location in competition and innovation, where clusters of companies, suppliers, and related institutions contribute to productivity and competitive advantage (Porter, 1998). These localized clusters provide three core mechanisms through which economies are achieved: labor market pooling, input sharing, and knowledge spillovers (Marshall, 1920; Rosenthal and Strange, 2004).

The first static externality is thus labor market pooling, involving the concentration of a specialized workforce in a specific region. This phenomenon provides a beneficial platform for employers and potential employees, reducing costs by providing access to a concentrated pool of skills and expertise tailored to their industry's needs (Rosenthal and Strange, 2004). Another static benefit is the sharing of inputs, where spatially proximate firms in the same industry can share resources, leading to internal increasing returns to scale and cost-effective operations (Duranton and Puga, 2004). Additionally, spatial proximity to suppliers and customers facilitates joint innovation along the value chain. This closeness reduces transport costs and inventories, further enhancing efficiency (Storper and Venables, 2004). However, a higher concentration of similar firms in the same region can also result in intensified competition (Porter, 1998; Basile et al., 2017). The third and dynamic source of localization economies arises from knowledge spillovers. These spillovers are particularly significant in localized industries, fostering knowledge exchange among employees (Rosenthal and Strange, 2004). According to the Knowledge Spillover Theory of Entrepreneurship (KSTE), these spillovers from universities, firms, and other sources significantly influence entrepreneurship and firm performance (Ghio et al., 2015; Cassia and Colembelli, 2008). Knowledge spillovers facilitate the development of innovative products and services, increased efficiency, and enhanced market competitiveness. Ghio et al. (2015) further discuss that knowledge spillovers can enable startups to overcome resource constraints and expand their networks, leading to increased opportunities for partnerships, collaborations, and investments. Hence, these spillovers can result in the adoption of new technology, increased product and process innovation, and enhanced marketing capabilities.

2.2.2 DIVERSIFICATION EXTERNALITIES

Diversification externalities, unlike localization externalities, result from the coexistence of various industries within a local economy (Basile et al., 2017). This concept, rooted in Jacobs' works (1969), highlights that this diversity creates opportunities for knowledge integration across industries (Glaeser et al., 1992). Solutions implemented in one industry can then be readily adapted to address challenges in other industries. Moreover, diversification provides firms with more stable demand conditions and a wide range of local input substitutes, reducing uncertainty (Neffke et al., 2011).

2.2.3 URBANIZATION EXTERNALITIES

Urbanization externalities are the benefits enjoyed by firms located in large cities. These cities provide access to expansive markets, both locally and as hubs in international infrastructure networks. Additionally, they offer access to highly educated employees, strong research and development centers, and efficient business services. Basile et al. (2017) suggest that urbanization economies, represented by the overall density of local economic activity, provide businesses with diverse local public services and increased local demand, ultimately enhancing firm performance (Ciccone and Hall, 1996). Nevertheless, high urban density can lead to disadvantages such as increased land prices, congestion, and higher labor costs, which may outweigh the benefits (Duranton and Puga, 2004; Bosma et al., 2008). The relationship between urban density and its associated externalities may exhibit nonlinearity, with positive agglomeration economies dominating up until a certain threshold, beyond which congestion effects become prominent (Basile et al., 2013).

2.3 AGGLOMERATION ECONOMIES ON STARTUP PERFORMANCE

Urban areas provide startups with valuable resources and opportunities for growth and development (Minai and Lucky, 2011; Schutjens and Wever, 2005). Specifically, the proximity to high-tech firms, customers, and suppliers, along with access to a skilled labor force, make urban areas attractive for startups (Frenkel, 2012; van Geenhuizen and Reyes-Gonzalez, 2007). Furthermore, startups located in urban areas can benefit from the demographic structure and institutional environment, easier access to financial resources, and closer market proximity (Cavallo, 2020; Tunberg, 2014; Felsenstein and Fleischer, 2002). The presence of an entrepreneurial ecosystem, as described by Bosma et al. (2008), plays a critical role in startup location choices, with urban areas often offering better-equipped ecosystems. Such ecosystems thus encompass various forms of capital, networks, knowledge transfers, and competition. Urban areas are known for providing these advantages, including a large pool of skilled labor, good transport links, and access to capital. Business incubators and science parks, as highlighted by Qian et al. (2016), increasingly support startup development by providing resources and networks. These initiatives are often situated in urban areas and are frequently associated with universities, which are typically located in larger cities (Oakey, 2007; Cavallo, 2020).

2.3.1 URBANIZATION AND VENTURE CAPITAL

Access to venture capital significantly impacts startup success, as demonstrated by Jeong et al. (2020). They highlight that young firms often lack financial resources and intangible assets, such as experience and knowledge, which venture capitalists (VCs) provide to foster startup growth. However, Catalini et al. (2017) suggest that the role of VCs in promoting startup growth has diminished over time, as startups increasingly turn to alternative funding sources like crowdfunding and incubators. Chen et al. (2009) reveal that the venture capital distribution is unevenly spread, with most of it concentrated in metropolitan cities. Nonetheless, Florida and Mellander (2017) find that VC-financed innovation is experiencing a changing geographical distribution in the US due to the availability of skilled workers elsewhere and higher transaction costs in traditional startup clusters.

Fritsch and Schilder (2008) suggest that spatial proximity is important for VC investment, especially for early-stage and high-tech firms. They find that startups located closer to venture capital firms have a higher likelihood of securing funding (Felsenstein and Fleischer, 2002). This proximity influences the search for investment opportunities, as face-to-face contact is necessary for monitoring and supervision, leading to increased transaction costs with greater geographical distance (Fritsch and Schilder, 2008; Zook, 2002).

However, a later study by Fritsch and Schilder (2012) challenges their earlier findings, as it reveals that the average distance between investors and investments exceeds 232 kilometers, due to "VC syndication", where risks, capital, and expertise are shared among multiple financiers, reduces the cost of investment activities. Additionally, the increased use of information technologies, accelerated by the Covid-19 pandemic, has facilitated efficient communication between investors and startups, reducing the importance of spatial proximity (Almeida, 2020). This has also enhanced market access and transport infrastructure, allowing firms to thrive beyond major cities (Bel and Fageda, 2008).

2.4 GEOGRAPHICAL CENTRALITY

Burgers and Meijers (2011) argue that geographical centrality, a concept rooted in economic geography and spatial analysis, refers to a location's relative significance within a larger spatial context. This concept is used to reveal patterns of human settlement, economic activity, and other forms of spatial interaction (Eaton and Lipsey, 1982; Mulligan, 1984).

Geographical centrality is intrinsically linked to theories of urban development and spatial economics, including the Central Place Theory (CPT). The CPT asserts that degrees of centrality vary across the economic landscape, a phenomenon thus termed "geographical centrality" (Christaller, 1933). The theory posits that a location's relative centrality significantly impacts its accessibility to surrounding markets. Central places are hierarchically organized, with larger ones serving more extensive areas and providing a range of goods and services to broader markets (Eaton and Lipsey, 1982; Mulligan, 1984). This hierarchy influences the clustering of economic activities in certain regions due to the benefits of co-location (Mulligan, 1984). Parr (2017) defines a "central place" as an urban center predominantly engaged in exporting goods to its surrounding market areas, which may be primarily rural. Such centers serve both local and external demands, showing exogenous characteristics. These centers also maintain endogenous, or center-serving, characteristics, mainly fulfilling local needs (Parr, 2017).

Bourdeau-Lepage et al. (2009) stress that a center is an agglomerated area for population, employment, economic activities, wealth creation, knowledge, innovation, and power. However, centrality extends beyond economic entity agglomeration; it also encompasses accessibility and interactions enabled by transport and communication networks. These networks provide interactions between the center and other places (Bourdeau-Lepage et al., 2009). According to Mulligan (1984), economic activities thus cluster in certain regions due to the benefits of co-location, such as access to labor, capital, and knowledge networks.

Despite the traditional model of hierarchy and mono-centricity, more recent perspectives suggest that centrality is becoming more dispersed, and fragmented due to globalization, technological advancements, and metropolisation (Burger and Meijers, 2011). To explain these dynamics, the concept of polycentricity has emerged, suggesting that urban systems can be organized around multiple centers of approximately equal importance (Parr, 2004).

3. CONCEPTUAL MODEL

In this chapter the conceptual model is developed to investigate the relationships between agglomeration economies and geographic centrality on startup performance. This model guides the empirical investigation of the factors that drive startup performance and addresses the research questions. See figure 1 for the conceptual model.

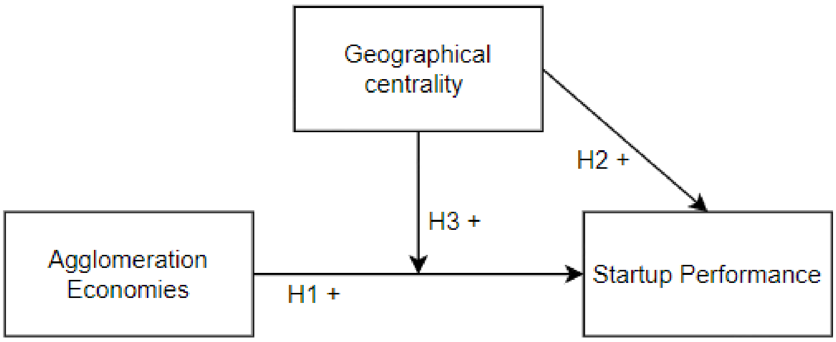


Figure 1: conceptual model of the relationships with hypotheses

3.1 HYPOTHESES

H1: Agglomeration economies have a positive effect on startup performance.

To support this hypothesis, we consider the externalities associated with agglomeration economies: localization, diversification, and urbanization externalities, which offer potential benefits and costs for businesses located in proximity to other economic actors (Neffke et al., 2011). Existing literature provides substantial evidence supporting the positive relationship between agglomeration economies and startup performance. Studies consistently find that startups in urban areas outperform those in rural areas in terms of profitability, employment growth, and innovation (Thwaites and Wynarczyk, 1996; Keeble and Walker, 1994). Urban areas, with their high concentration of high-tech firms, universities, and business incubators, foster entrepreneurial ecosystems that provide startups with valuable resources, networks, and knowledge transfers (Bosma et al., 2008; Qian et al., 2016).

Startups in urban areas also benefit from proximity to universities and business incubators, which enhance performance through resource access, networks, and knowledge institutions (Oerlemans and Meeus, 2005; Qian et al., 2016). Urban areas further offer better access to venture capital, a crucial factor for startup growth and success (Jeong et al., 2020; Fritsch and Schilder, 2008). Additionally, knowledge spillovers facilitated by universities, firms, and other sources of knowledge in urban areas positively influence entrepreneurship and firm performance (Wennberg et al., 2011; Ghio et al., 2015). However, it is important to acknowledge potential disadvantages of urban agglomerations, including congestion, competition, and higher costs such as land prices and labor (Duranton and Puga, 2004; Bosma et al., 2008; Li et al., 2016). Nonetheless, some studies suggest that startups in rural areas can achieve better performance outcomes due to factors such as lower costs, less competition, and access to established professional networks (Dahl and Sorensen, 2012; Yu et al., 2011). Despite the conflicting findings, the substantial evidence from the literature make it reasonable to assume that agglomeration economies positively impacts startup performance (H1).

H2: The geographical centrality of the startup's location has a positive effect on startup performance.

Drawing from the Central Places Theory (Christaller, 1933; Eaton and Lipsey, 1982) and Agglomeration Theory (Mulligan, 1984), the assumption that startups in geographically central regions have better performance outcomes than those in non-central or peripheral regions carries significant theoretical weight. These theories emphasize the impact of consumer behavior, market competition, and transportation costs on economic activity distribution. Geographically central regions possess high accessibility facilitated by transport and communication networks, which shape interactions between the center and other places (Bourdeau-Lepage et al., 2009). Mulligan (1984) suggests that activities cluster in specific regions due to co-location benefits, including access to essential components like labor, capital, and knowledge networks (Frenkel, 2012).

While Gallagher and Miller (1991) found that central areas in the UK had a greater number of fast-growing companies compared to peripheral areas, Stam (2005) discovered no region-based differences in fast-growing firms in their study focused on the Netherlands. This finding was attributed to the Dutch city system's structure, characterized by a single polycentric urban field. Similarly, Anderson et al. (2005) found no statistically significant difference in employment or revenue changes among central, accessible-central, and remote-periphery firms. These findings suggest that geographic centrality may not have as significant an impact on startup performance as originally theorized.

However, despite the results of Gallagher and Miller (1991) and Anderson et al. (2005), the theoretical underpinnings from the Central Places Theory and Agglomeration Economies Theory provide strong motivation to expect a positive relationship (H2) and warrant further investigation into the hypothesis that geographical centrality positively influences startup performance.

H3: The relationship between agglomeration economies and startup performance is moderated by the geographical centrality of the startup's location.

The direct effects between agglomeration economies and geographical centrality on startup performance have been extensively studied. But the joint influence of these factors on startup performance remained largely understudied. Meccheri and Pelloni (2006) observed that rural areas in the "periphery" differ from those in the "center." Startups located in central areas, both urban and rural, tend to have access to a greater variety of resources, networks, and markets compared to their more peripheral counterparts. Thus, it is reasonable to assume that startups in centrally located areas, regardless of their urban or rural nature, exhibit better performance outcomes. Consequently, geographical centrality may amplify the impact of a startup's location on its performance.

Urban startups benefit from urbanization externalities, such as knowledge spillovers, resource sharing, and diverse labor pools. And, startups located in centrally placed urban areas may experience even greater advantages due to access to larger markets, robust transportation networks, and enhanced resource availability. However, rural startups, often viewed as disadvantaged due to resource constraints, may outperform their urban counterparts when situated in geographically central areas. This can be attributed to avoiding agglomeration diseconomies prevalent in cities while leveraging the benefits of central regions, such as larger markets and improved infrastructure (Li et al., 2016; Bourdeau-Lepage et al., 2009).

Supporting this argument, studies by Reynolds et al. (1994) and Storey (1994), as cited by Schutjens and Wever (2005), have found faster growth rates for rural startups located in more central areas compared to urban startups. Building on the extensive literature, it is expected that geographic centrality amplifies the benefits of agglomeration economies. This prompts a more nuanced understanding of location-dependent factors in startup performance and encourages exploration of potential synergistic effects between agglomeration economies and geographical centrality (H3).

4. METHODS

This chapter provides an overview of the research methods employed, including the study design, sampling techniques, and data collection methods. It also discusses the measurements used, data analysis techniques (including descriptive statistics), and the assumptions underlying the regression analysis.

4.1 RESEARCH DESIGN

Given the research objective of examining the impact of agglomeration economies and geographical centrality on startup performance, a quantitative research approach is deemed appropriate as it is well-suited to address such questions (Goertzen, 2017). This study adopts a quantitative and descriptive research design, utilizing existing secondary data. Secondary data is preferred in this case due to its availability in large and diverse datasets, which are valuable for this research and may be more challenging to obtain through smaller primary sources (Devine, 2003).

4.1.1 RESEARCH ETHICS

Stringent ethical standards are upheld in this research, despite relying on secondary data and the absence of direct interaction with human participants. All data used in the study is publicly available and respects the privacy of the startups, ensuring that no personally identifiable information is used or disclosed. Transparency and accountability are prioritized, maintaining the integrity of the data (APA, 2017). Proper referencing is employed for all data sources, and every step in the data cleaning, preparation, and analysis processes is thoroughly reported to ensure reproducibility.

Bias is carefully avoided in the selection of startups, solely based on the availability of required data and without personal preferences influencing the choices. The research emphasizes adherence to principles of validity and reliability, ensuring that the chosen methods and statistical tests effectively address the research question and yield reliable results. Intellectual property rights are respected, with diligent acknowledgment and referencing of all information sources, including databases and prior research studies (APA, 2017). Lastly, the research aims to contribute positively to societal discourse surrounding urban planning and startup ecosystem development, with an objective and transparent reporting of the results.

4.2 SAMPLING AND DATA COLLECTION

Crunchbase is widely recognized and commonly used as a data source for startup research (Dalle et al., 2017). Previous studies by Block and Sandner (2009) and Kaminski et al. (2019) have also utilized Crunchbase for their research on startups. The availability of information on

Crunchbase, including details on the number of employees, industry, headquarters location, funding, and estimated revenue, makes it a valuable resource for this study. The research utilized an existing database from 2020, encompassing data on nearly one million startups worldwide.

The sample consists of independent startups in the Netherlands, chosen due to the vibrant startup ecosystem and high number of startups per capita across various industries (Stam, 2014). This study will be focusing on independent startups as it provides clarity in insights, because acquired startups could introduce complexity and potential distortion, especially in obtaining reliable and complete funding data for privately held companies. However, it is important to note that the findings may be limited to the context of the Netherlands and may not be easily extrapolated to other countries. Purposive sampling is used in this study, as specific information about these particular startups is needed (Etikan et al., 2016).

4.3 VARIABLES AND MEASUREMENTS

Dependent variable. Measuring startup performance requires considering various indicators. The amount of funding raised and estimated revenue are commonly used indicators of viability and success in startup performance research (van Rijnsoever et al., 2017; Thwaites and Wynarczyk, 1996; Dahl and Sorensen, 2012). Therefore, these two indicators will be used to measure startup performance. It is important to note that this study uses estimated revenue as indicator due to the lack of disclosure for private companies about this type of company information (Eveleens et al., 2017).

Independent variable. Initially, the intended measure of agglomeration economies was address density as a continuous variable. However, unexpected challenges arose, leading to a methodological shift. The five-level classification of urban density proposed by den Dulk et al. (1992) was chosen, which is recognized by the Central Bureau of Statistics of the Netherlands (see table 1). This adjustment allowed for navigating the complexities encountered. Appendix 1 provides a detailed account of these issues and the journey towards the solution. It is advisable to review this appendix for a comprehensive understanding of the changes. For the urban densities per municipality, please refer to table appendix 2b. This variable is dummified for the interpretation, where extremely urbanized is set as reference category.

Table 1. Classification Urban Density in the Netherlands (den Dulk, 1992)

Extremely urbanized	2500 or more addresses per km2 on average
Strongly urbanized	1500-2500 addresses per km2 on average
Moderately urbanized	1000-1500 addresses per km2 on average
Hardly urbanized	500-1000 addresses per km2 on average
Not urbanized	less than 500 addresses per km2 on average

Moderator: Geographic centrality.

Geographic centrality is measured using the COROP-regions, which divide the Netherlands into three zones based on the number of jobs reachable within a 50 km radius. This classification, proposed by the Central Bureau of Statistics (CBS, n.d.), includes the most Central Zone (Randstad), the Intermediate Zone, and the National Periphery. The variable is transformed into dummy variables, as shown in table appendix 2a.

Figure 2: geographical centrality in the Netherlands. Legenda: brown is Central Zone, pink is Intermediate Zone, yellow is National Periphery

The startup's location is determined by its municipality, which corresponds to its geographic centrality (see figure 2 for visualization of the areas). It is important to note that the use of this measurement has a limitation: the boundaries of the areas may not align perfectly with natural or economic regions. As a result, there is a possibility of distortions or oversimplifications when assessing centrality. It may not fully account for specific localities or influential urban centers that contribute to economic activity and innovation. For instance, Eindhoven serves as a significant startup hub but falls within the Intermediate Zone.

Control variables. Control variables were included in the analysis to prevent potential distortions in the results. The control variables considered were: number of funding rounds, number of investors, and number of employees. Funding rounds and number of investors are ratio variables, representing the frequency of funding raised and the number of investors in each round, respectively. The number of employees, although gathered in ranges, is treated as a continuous variable by using the mean value of the range.

4.4 DATACLEANING AND PREPARATION

The dataset was refined by filtering Dutch startups and selecting those with available information on Total Funding in the 2020 dataset, resulting in an initial selection of 1022 startups located in the Netherlands. Three instances of duplicate entries were identified and removed (refer to table 2). Outliers were identified through visual inspection and initial regression analyses, revealing significant outliers and a low R-squared value, particularly among older startups. Given that Crunchbase was established in 2007, concerns were raised regarding the accuracy and completeness of data for firms founded prior to this year. Consequently, 159 startups established before 2007 were excluded from the dataset.

Although an initial filtering process was applied, 22 acquired startups were still found in the database. To address persistent discrepancies, a comprehensive data verification process was implemented, cross-verifying information for each startup using multiple databases such as Crunchbase, Tracxn, Dealroom.co, and LinkedIn (accessed in May 2023). This meticulous data auditing resulted in the removal of 72 non-operational firms, 9 non-Dutch startups, 1 non-profit organization, and 28 startups with undisclosed funding amounts. Inaccuracies related to the Number of Employees, Funding Rounds, and Investors were also rectified. The final sample used in the analysis consisted of the most recent information as of 2023, including data from 727 Dutch startups. This rigorous data cleaning and preparation process ensured the reliability of the dataset.

Table 1. Overview of deleted cases

Reason of deletion	*N*
Operating status: closed	72
Acquired	22
Undisclosed funding amount	28
Non-Dutch startup	9
Not-Profit	1
Double	3
Founded before 2007	159

4.5 DATA ANALYSIS

To investigate the research questions, ordinary least squares (OLS) regression analysis is employed. OLS regression analysis, as discussed by Hair et al. (2018), offers several strengths. It enables modeling of complex relationships involving multiple independent variables, allows for control of confounding variables, and provides estimates of the relationship's strength and direction.

4.6 DESCRIPTIVE STATISTICS (TABLE 3 AND TABLE 4)

For several variables in this analysis, a log transformation was applied to address skewed distributions, in turn, improving the likelihood of meeting the assumptions of OLS-regression analysis. This transformation is common in statistical analysis and involves taking the logarithm of each data point in a variable, which can help to minimize the impact of extreme values and transform into a more symmetric one (Hair et al., 2018).

- **Total Funding.** This variable ranges from 0 to 1.100 million dollars across 698 observations, with 29 (3,99%) missing values. The mean value is 11,89 million dollars, and the standard deviation is 49,64 million dollars, indicating a moderate dispersion around the mean. The variable exhibits positive skewness (15,916) and kurtosis (332,118), indicating a heavily tailed, right-skewed, and highly peaked distribution. To address this, a log transformation was performed, resulting in a significant improvement of the distribution. In line with the approach suggested by Hair et al. (2018), cases with missing data for dependent variables were deleted to prevent artificial inflation of relationships
- **Total Revenue.** Out of the total 565 observations in this variable, 162 (22,28%) have missing values, and as mentioned earlier, these cases with missing values were deleted. It is important to note that Total Revenue is considered a secondary variable in this research and is already estimated. The available data for Total Revenue ranges from 0,08 to 409,8 million dollars. The mean value is approximately 11,41 million dollars, and there is a wide dispersion indicated by the standard deviation of 28,04. The distribution exhibits high skewness (8,230) and kurtosis (90,788), indicating significant non-normality. To address this, a log transformation (as described earlier) was applied.
- **Number of Employees.** The number of employees in the dataset ranges from 5,5 to 750,5, with a mean value of 38,61. The standard deviation of 70,24 indicates significant variability around the mean. The distribution of the variable exhibits positive skewness (5,191) and kurtosis (36,595), indicating a highly skewed and peaked distribution. As mentioned earlier, a log transformation was applied to improve the distribution significantly.
- **Number of Investors.** This variable consists of 574 observations, with 152 (20,94%) missing values. The mean number of investors is 3,24, and the standard deviation is relatively low at 3,19. The distribution exhibits positive skewness (2,486) and kurtosis (8,280), indicating a right-skewed and peaked distribution. A log transformation was applied, resulting in a significant improvement in the distribution. According to Hair et al. (2018), missing values above 10% cannot be ignored. Additionally, the deletion of a variable can have reduced negative effects if it is highly correlated with another independent variable and represents the original variable's intent. Given the strong correlation between the Number of Investors and Number of Funding Rounds (refer to

table appendix 3a), the decision was made to remove the Number of Investors variable from the analysis.

- **Number of Funding Rounds.** This variable ranges from 1 to 14 rounds, with an average of approximately 2,43 rounds and a standard deviation of 1,82. The distribution exhibits positive skewness (1,846) and kurtosis (4,525), indicating a right-skewed and peaked distribution. To improve the distribution, a log transformation was applied, resulting in a significant improvement.
- **Urban Density.** The data presents the distribution of urban density across the National Periphery, Intermediate Zone, and Central Zone. It is noteworthy that the majority of areas in the National Periphery (58,9%) are classified as 'Hardly Urbanized', while the Intermediate Zone is predominantly 'Moderately Urbanized' (42,7%). The Central Zone stands out with an overwhelming majority (65,6%) of areas categorized as 'Extremely Urbanized'. When considering all centrality levels, 'Extremely Urbanized' areas constitute the largest category overall (51,9%). This skewed distribution towards extreme urbanization, particularly in the Central Zone, represents a notable finding. For a detailed breakdown of the distribution of categories within each zone and overall, please refer to the provided table 4.
- **Geographical Centrality.** Most startups, 575 (79,2%), are based in the most Central Zone of The Netherlands. Meanwhile, 96 (13,2%), are located in the Intermediate Zone, and a mere 56 (7,6%) in the National Periphery.

Table 3. Descriptive statistics

	Variable [a]	N	Min	Max	Mean	Std. Dev	Skewness	Std. Error	Kurtosis	Std. Error
1.	Total Funding [b]	699	0	1.100	11,89	49,64	15,916	0,093	332,118	0,185
	~Log transformation	-	-	-	-	-	0,046	0,093	-0.304	0,185
2.	Total Revenue [c]	565	0,08	409,80	11,41	28,04	8,230	0,103	90,788	0,205
	~Log transformation	-	-	-	-	-	0,658	0,103	1,436	0,205
3.	Number of Employees	727	5,5	750,5	38,61	70,24	5,191	0,091	36,595	0,181
	~Log transformation	-	-	-	-	-	0,483	0,091	-0,370	0,181
4.	Number of Investors	574	1	24	3,24	3,19	2,486	0,102	8,280	0,204
	~Log transformation	-	-	-	-	-	0,577	0,102	-0,578	0,204
5.	Number of Funding Rounds	727	1	14	2,43	1,82	1,846	0,091	4,525	0,181
	~Log transformation	-	-	-	-	-	0,490	0,091	-0,858	0,181

[a]n = 727 startups; [bc]in million dollars

Table 4. Frequencies of Urban Density and Geographic Centrality

	National Periphery	Intermediate Zone	Central Zone	**Total**
Not Urbanized	13 (23,2%)	16 (16,7%)	24 (4,2%)	53 (7,3%)
Hardly Urbanized	33 (58,9%)	24 (25,0%)	13 (2,3%)	70 (9,6%)
Moderately Urbanized	10 (17,9%)	41 (42,7%)	67 (11,7%)	118 (16,2%)
Strongly Urbanized	0 (0%)	15 (15,6%)	94 (16,3%)	109 (15,0%)
Extremely Urbanized	0 (0%)	0% (0%)	377 (65,6%)	377 (51,9%)
Total	56 (7,7%)	96 (13,2%)	575 (79,1%	**727 (100%)**

4.7 ASSUMPTIONS TESTING

The multicollinearity diagnostics for this research are thoroughly presented in Appendix 3. This section includes a comprehensive analysis of each variable in the model, examining tolerance and Variance Inflation Factor (VIF). The control variables exhibit no serious multicollinearity, as indicated by a minimum tolerance of 0,811 and a VIF of 1,233. However, certain direct effects surpass the accepted threshold for multicollinearity (a VIF of 5,8 and corresponding tolerance of 0,172), requiring careful interpretation of the interaction terms. Interaction effects involving variables with no intersecting data or zero contribution to the model were excluded.

Linearity diagnostics, detailed in Appendix 4, demonstrate that linear regression models generally yield more interpretable results for these factors, despite the presence of statistically significant polynomials. Appendix 5 validates the assumption of normality of residuals for models predicting total funding, while the models predicting total revenue exhibit slight violations of this assumption. Lastly, homoscedasticity is assessed in Appendix 6. Overall, the model assumptions are mostly met, but it is important to be cautious in interpreting the results due to the slight violations.

5. REGRESSION ANALYSIS RESULTS (TABLE 5)

Table 5. Model Summary and Regression Coefficients

	(Log) Total Funding			(Log) Total Revenue		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
(Constant)	-3,612 **** (-25,874)	-3,473 **** (22,286)	-3,486 **** (-22.523)	-0,259 *** (-2,626)	-0,173 (-1,545)	-0,162 **** (-1,439)
(Log) Number of Employees	1,130 **** (23,254)	1,113 **** (22,957)	1,121 **** (23,254)	0,596 **** (17,937)	0,586 **** (17,476)	0,582 **** (17,276)
(Log) Number of Funding Rounds	1,008 **** (11,945)	0,993 **** (11,798)	0,978 **** (11,685)	0,110 * (1,891)	0,096 (1,628)	0,096 (1,630)
Extremely Urbanized Δ Strongly Urbanized		-0,184 (1,237)	-0,130 (0,840)		-0,015 (0,143)	0,001 (-0,005)
Extremely Urbanized Δ Moderately Urbanized		0,160 (-0,997)	0,265 (1,479)		-0,037 (0,317)	0,018 (-0,133)
Extremely Urbanized Δ Hardly Urbanized		0,458 ** (-2,024)	0,655 * (1,716)		0,140 (-0,887)	0,393 (-1,481)
Extremely Urbanized Δ Not Urbanized		0,825 **** (3,789)	0,914 *** (-3,287)		0,320 ** (-2,020)	0,278 (-1,367)
Central Zone Δ Intermediate Zone		0,060 (-0,340)	0,002 (-0,009)		0,077 (-0,613)	-0,082 (0,431)
Central Zone Δ National Periphery		-0,447 * (1,886)	-0,336 (0,766)		0,094 (-0,550)	-0,131 (0,424)
Not Urbanized * National Periphery			0,323 (0,513)			-0,070 (-0,153)
Moderately Urbanized * National Periphery			1,071 * (1,847)			-0,312 (-0,763)
Not Urbanized * Intermediate Zone			0,061 (0,123)			-0,415 (-1,143)
Strongly Urbanized * Intermediate Zone			0,328 (0,732)			-0,048 (-0,146)
Hardly Urbanized * Central Zone			1,247 ** (2,395)			0,495 (1,372)
Diagnostics						
R	0,782	0,790	0,796	0,655	0,663	0,666
Adjusted R^2	0,610	0,625	0,626	0,427	0,431	0,443
N	699	699	699	565	565	565
Degrees of freedom	2	8	13	2	8	13
F – value	546,769 ****	143,447 ***	91,030 ****	211,475 ****	54,435 ****	33,706 ****

* p ≤ 0,10, **p ≤ 0,05, ***p ≤ 0,01, ****p ≤ 0,001; (Constant) = (Ln) Total Funding/Revenue in million dollars; accompanying t-values in parentheses; Δ = in comparison of

5.1 MODEL SUMMARY

This section presents the results of the hierarchical regression models used to examine the hypotheses. The analysis began with a base model (Model 1) comprising control variables. It was followed by the inclusion of direct effects of predictors (Model 2) and interaction effects (Model 3). The OLS regression analysis of Total Funding and Total Revenue revealed that Models 1 and 2 provide the most reliable insights.

5.2 MODEL SUMMARY AND CONTROL VARIABLES

For Model 1, with (Log) Total Funding as the dependent variable, the R^2 value of 0,610 indicates that approximately 61% of the variation in Total Funding can be explained by the control variables, namely (Log) Number of Employees and (Log) Number of Funding Rounds. In the case of (Log) Total Revenue, the control variables explain 42,7% of the variation. These models demonstrate a significant fit to the data, as evidenced by the significant F-statistics ($F(2, 698) = 546{,}769$, $p \leq 0{,}001$ for Total Funding, and $F(2, 564) = 211{,}475$, $p \leq 0{,}001$ for Total Revenue).

In Model 2, after incorporating the direct effects, the explanatory power increased to 62,5% for Total Funding and 43,1% for Total Revenue. Compared to Model 1, this indicates that the urban density and geographical centrality variables account for an additional 1,5% and 0,4% of the variance in Total Funding and Total Revenue, respectively.

For Model 3, the interaction effects were introduced, resulting in a further increase in the R^2 value to 62,6% for Total Funding and 44,3% for Total Revenue. This suggests that the interaction terms explain an additional 0,1% and 1,2% of the variance in Total Funding and Total Revenue, respectively.

5.2.1 DIRECT EFFECTS: HYPOTHESES 1 AND 2

H1: Agglomeration economies have a positive effect on startup performance. The impact of agglomeration economies (urban density) on startup performance was assessed, revealing varying effects. Please refer to table 1 for the dummified categories. Startups in Extremely Urbanized areas showed no statistically significant difference in Total Funding compared to Strongly or Moderately Urbanized areas (coefficients of -0,184 and 0,160, respectively; $p \geq 0{,}10$). However, a notable shift occurred when comparing Extremely Urbanized areas to Hardly Urbanized areas. Startups in Extremely Urbanized areas significantly outperformed their counterparts in terms of Total Funding, with a coefficient of 0,458 ($p \leq 0{,}05$). This translates to approximately 58,1% more funding received by startups in Extremely

Urbanized areas compared to Hardly Urbanized areas ([exp(0,458) - 1] * 100%). A more pronounced trend emerged when examining startups in Extremely Urbanized areas versus Not Urbanized areas. The coefficient of 0,825 ($p \leq 0,001$) strongly indicates that startups in Extremely Urbanized areas perform significantly better than those in Not Urbanized areas, receiving approximately 128,7% more funding ([exp(0,825) - 1] * 100%).

Regarding Total Revenue, the comparisons yielded diverse insights. Startups in Extremely Urbanized areas did not demonstrate a statistically significant difference in Total Revenue compared to Strongly, Moderately, or Hardly Urbanized areas (coefficients of -0,015, -0,037, and 0,140, respectively; $p \geq 0,10$). However, when comparing startups in Extremely Urbanized areas to Not Urbanized areas, the coefficient of 0,320 ($p \leq 0,05$) indicates that startups located in Extremely Urbanized areas generate significantly higher Total Revenue. This translates to approximately 37,7% more revenue for startups in Extremely Urbanized areas ([exp(0,320) - 1] * 100%).

In conclusion, the urban density of a startup's location has a positive effect on startup performance, particularly as the level of urbanization increases. Statistically significant differences were primarily observed when comparing startups in Extremely Urbanized areas to Hardly Urbanized areas, providing support for Hypothesis 1.

H2: The geographical centrality of the startup's location has a positive effect on startup performance. When analyzing the geographical centrality of a startup's location in relation to Total Funding, no significant difference is observed at the 5% level ($p \leq 0,05$). The same holds true for Total Revenue, as startups did not significantly differ based on their geographical centrality. Therefore, it can be concluded that geographical centrality does not seem to have an impact on Total Funding and Total Revenue for startups in the Netherlands. Consequently, Hypothesis 2 is not supported.

5.2.2 INTERACTION EFFECTS: HYPOTHESIS 3

H3: The relationship between agglomeration economies and startup performance is moderated by the geographical centrality of the startup's location. Multicollinearity in a regression model can complicate the analysis by hindering the identification of individual effects of predictors on the outcome variable. Due to high multicollinearity and strong correlation among predictor variables, several interaction terms were excluded from the analysis, impacting the reliability and interpretability of the model. The hypothesis proposed that geographic centrality would amplify the benefits of urbanization in terms of interaction effects.

A significant interaction effect was observed for Total Funding, while no significant effects were found for Total Revenue. Startups in the Central Zone categorized as Hardly Urbanized (1,247; $p \leq 0,05$) were more likely to secure greater amount of funding. This suggests that geographic centrality positively moderates the relationship between urban density and startup performance in the case of Hardly Urbanized areas in the Central Zone. Additionally, a positive effect was found for startups in Moderately Urbanized areas in the National Periphery. However, it is important to consider the potential issues associated with multicollinearity, which may compromise the precision of these estimates.

5.2.3 OVERVIEW OF RESULTS

While urban density had a varying effect on startup performance, providing partial support for Hypothesis 1, geographical centrality didn't significantly influence it, contradicting Hypothesis 2. Hypothesis 3, concerning the moderating effect of geographic centrality, showed mixed results due to multicollinearity (see table 6 for an overview of the results).

Table 6. Overview of results

Hypothesis	Relationship (path)	Proposed Effect	Observed Effect	Decision
H1	Agglomeration economies → Total funding	Positive	Positive	Supported
	Agglomeration economies → Total revenue	Positive	Non-significant	Not supported
H2	Geographic centrality → Total funding	Positive	Non-significant	Not supported
	Geographical centrality → Total revenue	Positive	Non-significant	Not supported
H3	Agglomeration economies * Geographical centrality → Total funding	Positive	Positive	Supported
	Agglomeration economies * Geographical centrality → Total revenue	Positive	Non-significant	Not supported

5.3 POST-HOC ROBUSTNESS TESTS

To establish the robustness and generalizability of the primary regression model and confirm the persistence of observed effects, additional post hoc tests were conducted, controlling for different aggregated industry groups (refer to appendix 7 for further explanation). The industry groups analyzed include Health Care, Retail and Commerce, Services, Technology, and Other, aiming to further investigate the relationships between agglomeration economies, geographic centrality, and startup performance. Due to small group sizes when considering interaction terms, these tests were exclusively performed for direct effects. The results (see table appendix 7b) indicate that across all industry groups, the logarithm of the number of employees and the number of funding rounds maintained their significantly positive relationships with total funding (p ≤ 0,001 and p ≤ 0,05). However, the effects of agglomeration economies and geographic centrality displayed more heterogeneity across industries (please refer to table 7 for an overview of the results).

Regarding agglomeration economies (see table appendix 7b), while the impact on startup performance was more pronounced in the Technology industry group compared to the initial regression model (with higher significance), the effects were much weaker in the remaining industry groups. Specifically, Health Care, Retail and Commerce, and Services showed no significant effect of agglomeration economies on startup performance. Only the Other industry group exhibited a significant but less pronounced positive effect. It is worth noting that the F-value for this particular industry group was low and not significant, which suggests that the model may not explain the variations well, potentially due to the high heterogeneity of industries within this group.

Moving on to geographic centrality, while no significant effect was found in the original model, significant effects were observed within specific industry groups (see table appendix 7b). Particularly notable is the highly significant positive effect of geographic centrality on the Services industry (for the Central Zone compared to the Intermediate Zone). Conversely, significantly negative effects of geographic centrality were identified for the Retail and Commerce and Technology industry groups.

Table 7. Overview of Post-Hoc Robustness Tests

Relationship (path)	**Industry group**	**Observed Effect initial OLS**	**Observed Effect Robustness Tests**
Agglomeration economies → Total funding	Health Care	Positive	Non-significant
	Retail and Commerce	Positive	Non-significant
	Services	Positive	Non-significant
	Technology	Positive	Stronger positive
	Other	Positive	Weaker significant
Geographic centrality → Total funding	Health Care	Non-significant	Non-significant
	Retail and Commerce	Non-significant	Negative
	Services	Non-significant	Positive
	Technology	Non-significant	Negative
	Other	Non-significant	Non-significant

6. DISCUSSION AND CONCLUSION

This discussion addresses the research question: "What are the effects of agglomeration economies and geographic centrality on startup performance, and does geographic centrality influence the relationship between agglomeration economies and startup performance?" This question reflects the ongoing debate in entrepreneurial literature regarding the importance of location in determining startup success (Minai and Lucky, 2011; Li et al., 2016; Frenkel, 2012; Oerlemans and Meeus, 2005). Drawing on theories such as Agglomeration Economies Theory and Central Places Theory, this research aimed to explore the role and impact of spatial factors on entrepreneurship, specifically focusing on total funding and total revenue. The discussion is structured around hypotheses, providing a comprehensive interpretation of the results, comparing them with existing literature, acknowledging limitations, discussing practical implications, and suggesting avenues for further research.

6.1 AGGLOMERATION ECONOMIES ON STARTUP PERFORMANCE

The results of Hypothesis 1 demonstrated a positive and significant effect of agglomeration economies (urban density). However, it is worth noting that a substantial gap was required to reveal a significant difference. Moreover, Model 2 had a very low explanatory power, suggesting that factors other than agglomeration economies likely have a strong influence on startup performance. The finding that startups in densely populated areas generally exhibit higher performance aligns with and expands upon previous scholarly work in this area. It supports the concept of agglomeration economies theory, which posits that the concentration of economic activities in an area leads to productivity gains and business success by providing increased access to capital, labor, and knowledge (Marshall, 1920; Duranton and Puga, 2004; Minai and Lucky, 2011; Frenkel, 2012). Additionally, the concentration of investors and financial institutions in such areas may facilitate access to crucial startup funding (Jeong et al., 2020; Fritsch and Schilder, 2008). Notably, although not statistically significant, the total funding decreased for extremely urbanized areas, which aligns with the notion that positive agglomeration economies dominate up to a certain threshold, beyond which congestion effects come into play (Basile et al., 2013).

It is worth noting that the impact of agglomeration economies is more pronounced in the Technology industry group, including AI, Data and Analytics, and Biotechnology. However, no significant effect of agglomeration economies on startup performance was found in the Health Care, Retail and Commerce, and Services industries. This highlights the importance of considering industry-specific characteristics when interpreting the influence of agglomeration economies.

In dense urban environments, the concentration of similar firms, universities, research institutions, and a highly skilled workforce fosters an environment conducive to knowledge spillovers (Oerlemans and Meeus, 2005; Qian et al., 2016; Cassia and Colembelli, 2008). The robustness tests indicate that the benefits of knowledge spillovers may be primarily concentrated within (high) technology startups, while being less pronounced in other industry groups. Technology startups operate in a fast-paced and rapidly evolving sector where access to and leverage of the latest knowledge can significantly impact performance (Wennberg et al., 2011; Ghio et al., 2015). Conversely, sectors such as Health Care, Retail and Commerce, and Services may rely more heavily on other factors for their performance.

The relationship between agglomeration economies and total revenue, as found in this study, was less pronounced than that of total funding. This suggests that a larger gap in urban densities was needed to reveal a difference in total revenue. It is important to note that the revenue of these startups is estimated, which may affect the accuracy of the information. While the literature often assumes that agglomeration-related advantages extend to overall firm performance, the findings of this research call for a more nuanced understanding. This difference in the impact of agglomeration economies on total revenue may be attributed to the unique characteristics of revenue generation, which are often influenced by factors such as organizational factors and the competitive landscape, rather than location-based factors alone (Hansen and Wernerfelt, 1989). Additionally, the concept of competition, as outlined in Porter's Cluster Theory (1998), may help explain this result. While agglomeration economies offer numerous benefits, it also entails a higher level of competition among businesses due to proximity and concentration, which can complicate the revenue generation process and offset some of the advantages of agglomeration economies

6.2 GEOGRAPHICAL CENTRALITY ON STARTUP PERFORMANCE

The results of this study did not support Hypothesis 2, which proposed a positive impact of geographical centrality on startup performance. Based on theoretical implications, it was expected that startups situated in central areas would have better access to resources such as labor, capital, and knowledge networks, thereby positively influencing their performance (Bourdeau-Lepage et al., 2009; Mulligan, 1984). However, the results revealed no significant differences in Total Funding and Total Revenue among startups based on their geographic centrality. This lack of support challenges the theoretical foundations derived from Central

Places Theory (Christaller, 1933; Eaton and Lipsey, 1982) and Agglomeration Economies Theory (Mulligan, 1984), both of which suggest that startups in central regions would likely experience better performance outcomes compared to those in non-central or peripheral regions.

The choice of the Netherlands as the sample for this study, with its three-level (central, intermediate, and national periphery) measurement, may explain the lack of significance in the effects. The results align more closely with studies by Stam (2005) and Anderson et al. (2005), which also found no statistically significant regional differences. This suggests that the impact of geographic centrality on startup performance may be less significant than traditionally theorized, reinforcing the idea that other organizational and environmental variables may have a greater influence on startup performance (Hansen and Wernerfelt, 1989).

Another explanation for these findings is the polycentric nature of Dutch cities, where the country itself can be seen as a large, single urban field (Stam, 2005), blurring the distinctions between different levels of centrality. The central zone (Randstad), comprising Utrecht, Zuid-Holland, and southern parts of Noord-Holland, serves as the economic hub of the Netherlands, housing major cities and a large number of accessible jobs (CBS, n.d.). However, considering the robust transportation and communication networks in the Netherlands, it is worth questioning whether the advantage of geographical centrality is as pronounced here as it might be in countries with less integrated infrastructure. Even startups located in the national periphery, with fewer immediately accessible jobs, seem to still be able to tap into opportunities within the Randstad.

Furthermore, the Netherlands possesses multiple thriving hubs, including those in the intermediate zone (e.g., Eindhoven) and the national periphery (e.g., Enschede and Groningen). These locations may challenge the traditional notion of geographical centrality, as startups in these areas can benefit from proximity to a significant number of jobs, as well as the presence of universities, incubators, and knowledge spillovers that foster startup development (Qian et al., 2016; Ghio et al., 2015).

Contrary to assumptions made in the traditional mono-centric model (Eaton and Lipsey, 1982; Christaller, 1933), the findings of this study align with theories that highlight the effects of polycentricity and global connectivity. This trend resonates with recent research, which suggests that centrality is becoming more dispersed and fragmented due to factors such as globalization, technological advancements, improved market access and transport infrastructure, and the rise of digital communication technologies (Burger and Meijers, 2011; Bel and Fageda, 2008; Almeida, 2020). Finally, government policies and incentives aimed at promoting economic development in peripheral areas may also contribute to attracting more funding to startups in these regions (Isenberg, 2010). Consequently, the advantages of geographical centrality may be diminished or neutralized, particularly in the context of the Netherlands.

Although this study did not find overall statistical support for Hypothesis 2, post hoc robustness tests revealed industry-specific influences of geographical centrality on startup performance. For instance, in the Services industry (e.g., Financial services, Advising, and Education), being in a central area was associated with a significant increase in secured funding. Surprisingly, startups in the retail and commerce and technology sectors exhibited an advantage in national peripheral locations. These findings challenge, to some extent, the assumptions made in the initial theoretical framework. Specifically, the observed patterns within specific industries (the benefits of geographical centrality for the services industry and the advantages of peripheral locations for the Technology and Retail and Commerce industries) cannot be readily explained using Agglomeration Economies Theory or Central Places Theory. Therefore, these results call for a more nuanced understanding of how geographical centrality interacts with industry-specific characteristics in shaping startup performance.

6.3 INTERACTION OF CENTRALITY AND AGGLOMERATION ECONOMIES

The analysis of Hypothesis 3, which proposed that the relationship between agglomeration economies and startup performance is moderated by the geographical centrality of the startup's location, yielded complex results. The presence of high multicollinearity in the data complicated the analysis, leading to the exclusion of several interaction terms and limiting the interpretability and reliability of the model. Therefore, caution must be exercised when drawing conclusions from these findings.

Partially supporting Hypothesis 3, a significant positive effect was observed for startups in Hardly Urbanized areas in the Central Zone in terms of Total Funding. This finding aligns with the theoretical frameworks of Central Places Theory and Agglomeration Economies Theory, which suggest that geographic centrality can amplify the benefits associated with urban density, thereby enhancing startup performance. This suggests that startups in less urbanized but central areas can leverage the agglomeration economies typically linked to central locations, such as access to larger markets, better infrastructure, and abundant resources. However, it is important to note that the explanatory power of this interaction is low, limiting its practical applicability.

Interestingly, this result is consistent with the observations of Reynolds et al. (1994) and Storey (1994), as cited by Schutjens and Wever (2005), who reported faster growth rates for rural startups located in more central areas compared to their urban counterparts. The geographical centrality of less urbanized areas may offset the limitations often associated with rural locations and harness the advantages associated with centrality, thereby contributing to enhanced startup performance.

6.4 STUDY LIMITATIONS

While this study offers valuable insights into the relationship between agglomeration economies, geographical centrality, and startup performance in the Netherlands, it is important to acknowledge its limitations. This section will discuss various factors, including multicollinearity issues and the choice of measures for startup performance, that should be taken into consideration.

6.4.1 SEVERE MULTICOLLINEARITY

When discussing the study findings, an important consideration arises regarding multicollinearity, a statistical phenomenon observed within the data (Hair et al., 2018). Multicollinearity was detected during the examination of interactions. Several factors tied to the study design may have contributed to this occurrence. Notably, the chosen sample of the Netherlands plays a role. As previously discussed, the geographical structure of the Netherlands leads to a high concentration of startups clustering in central areas (Florida and Mellander, 2016; Frenkel, 2012), where urban density is also higher. In fact, 65,6% of the startups are located in both central and extremely urbanized areas. Consequently, these two independent indicators, urban density and geographical centrality, tend to predict the same outcome.

The skewed distribution of startups across regions and different urban densities further complicates the interpretation of the findings. With smaller sample sizes in certain categories, it becomes more challenging to detect significant effects, increasing the risk of Type II error (incorrectly failing to reject a false null hypothesis) (Hair et al., 2018). Unfortunately, this cannot be resolved as the study specifically focused on all available startups in the Netherlands.

6.4.2 LIMITATIONS OF CHOSEN MEASUREMENTS AND VARIABLES

Another limitation of this study is that it assesses startup performance solely based on total funding and total revenue, which are commonly used indicators (van Rijnsoever et al., 2017; Thwaites and Wynarczyk, 1996; Dahl and Sorensen, 2012). However, these measures may not fully capture the multifaceted nature of startup performance. Other factors, such as employment growth, market share, gross profit, cost control, survival, or a successful exit, could also play significant roles in defining startup success, but they are not addressed in this research (Eveleens et al., 2017).

A third limitation is the reliance on estimates of revenue generation for startups, which may not accurately reflect their actual financial performance. Estimating revenue carries inherent limitations and potential inaccuracies, particularly for privately held companies (Eveleens et al., 2017). Furthermore, the assumptions for regression analysis were slightly violated when considering total revenue as the dependent variable. Therefore, the results should be interpreted with caution. Fourthly, there is a limitation regarding the representation of the number of employees in startups. The use of ranges, although necessary for the analysis, may introduce imprecision. Treating these ranges as continuous data also introduces uncertainty and reduces the specificity of the findings. A final limitation of the study is its primary focus on agglomeration economies (urban density) and geographical centrality as indicators, potentially overlooking other significant factors. Factors such as incubation participation, access to venture capital, and other organizational factors may also impact startup performance (Qian et al., 2011; Jeong et al., 2019; Hansen and Wernerfelt, 1989).

6.5 PRACTICAL IMPLICATIONS

The findings of this study offer several practical implications for entrepreneurs, investors, policy makers, and academia, providing an actionable insight into how location-based factors affect startup performance.

For Entrepreneurs and Investors. The results demonstrate that considering location at the founding of a startup is significantly beneficial, but as previously discussed, other factors play a more prominent role in predicting success (low explanatory power). The study reveals that agglomeration economies have a positive impact on funding levels, particularly for high technology startups. Therefore, if entrepreneurs are considering the benefits of agglomeration economies for their startup, substantial increases in urban densities are necessary to make significant differences. These findings are also relevant for investors, who should recognize that while agglomeration economies can be advantageous, they should primarily focus on alternative factors to determine the viability before investing in a startup.

For Policy Makers. The results can guide the development of more balanced regional policies that promote inclusive economic growth. Rather than concentrating resources solely in urban dense areas, policymakers should consider ways to strengthen agglomeration economies in less densely populated cities. This approach may involve initiatives such as improving infrastructure, enhancing access to funding, and fostering entrepreneurial ecosystems in these areas.

For Academia. This research offers a novel perspective to the academic literature on startup ecosystems, highlighting that agglomeration economies and geographical centrality advantages do not universally apply across industries. The variation observed in industry-specific outcomes underscores the need for refining the Agglomeration Theory and the Central Places Theory. By emphasizing the critical role of industry context, this study contributes to the ongoing academic discourse on the impact of location on startup performance.

6.6 FUTURE RESEARCH

The current study's findings, context, and limitations provide a fertile ground for future academic exploration in the entrepreneurship domain.

Refinement of Measures and Models. Future research could explore alternative or complementary measures for agglomeration economies and geographical centrality, considering the presence of multicollinearity and the limited predictive power of the models used in this study. For instance, determining the driving distance to the first large startup hub (e.g. Amsterdam, Utrecht, Den Haag, Rotterdam, and Eindhoven) could provide valuable insights into the effects of geographical centrality on startup performance.

Other Performance Metrics. In addition to measuring startup performance through total funding and total revenue, future research should consider incorporating additional indicators such as employment growth, market share, gross profit, survival rate, and successful exit. By doing so, a more holistic view of startup performance can be obtained (Eveleens et al., 2017).

Expanding the Geographical Context. The findings in the Dutch context highlight the importance of conducting comparative studies across different countries and regions to understand the generalizability of the relationship between agglomeration economies, geographical centrality, and startup performance. Such studies could also help mitigate multicollinearity. Combes (2000) argues that it is worthwhile to consider agglomeration economies separately across countries. Germany, for example, with its more multi-polycentric geography of economic activity (Parr, 2004), could be an interesting sample, potentially yielding different findings. Exploring these avenues can contribute to a richer and more nuanced understanding in the entrepreneurship literature.

Exploring Industry Dynamics. The robustness checks results indicate that a 'one size fits all' approach does not capture the nuances that exist between industries. Future research should therefore aim to further explore these industry-specific influences on startup performance, by looking at knowledge-intensity for example.

6.7 CONCLUSION

This study explored the relationships between agglomeration economies, geographical centrality, and startup performance in the entrepreneurial landscape of the Netherlands. The results support the positive impact of agglomeration economies, with startups in densely populated areas generally exhibiting higher performance. The findings align with the concept of Agglomeration Economies Theory, emphasizing the advantages of urban density in terms of access to capital, labor, and knowledge. However, the impact of agglomeration economies varies across industries, highlighting the importance of considering industry-specific characteristics.

Contrary to expectations, the study did not find significant support for the positive influence of geographical centrality on startup performance. The polycentric nature of Dutch cities and the presence of thriving hubs in intermediate and peripheral areas challenge the traditional mono-centric model. The results suggest that the advantages of geographical centrality may be diminished or neutralized in the context of the Netherlands, where startups can tap into opportunities across the country due to robust transportation and communication networks. However, industry-specific influences were observed, indicating that centrality can still play a role in certain industries.

These findings have practical implications for entrepreneurs, investors, policy makers, and academia. Entrepreneurs and investors should consider the benefits of agglomeration economies while also focusing on other factors influencing startup viability. Policy makers should develop balanced regional policies that promote inclusive economic growth and strengthen agglomeration economies in less densely populated areas.

The limitations of this study, including multicollinearity issues and limited measures of startup performance, signal for future research focusing on exploring alternative measures and models. Considering for example additional performance metrics, and expand the geographical context to comparative studies across different countries and regions and industries. These results emphasize that geographical factors still play a crucial role in entrepreneurship but in nuanced and industry-specific ways. But the concluding remark is that the success of a venture is more determined by the 'how' and the 'what' instead of the ‘where’.

REFERENCES

Almeida, F., Duarte Santos, J., & Augusto Monteiro, J. (2020). The Challenges and Opportunities in the Digitalization of Companies in a Post-COVID-19 World. *IEEE Engineering Management Review*, *48*(3), 97–103. https://doi.org/10.1109/EMR.2020.3013206

Anderson, D., Tyler, P., & McCallion, T. (2005). Developing the rural dimension of business-support policy. *Environment and Planning C: Government and Policy*, *23*(4), 519–536. https://doi.org/10.1068/c44m

Basile, R., Donati, C., & Pittiglio, R. (2013). Industry structure and employment growth: Evidence from semiparametric geoadditive models. *Région et Développement*, *38*, 121–160.

Basile, R., Pittiglio, R., & Reganati, F. (2017). Do agglomeration externalities affect firm survival? *Regional Studies*, *51*(4), 548–562. https://doi.org/10.1080/00343404.2015.1114175

Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, *38*(2), 318–337. https://doi.org/10.1016/j.respol.2008.11.010

Bel, G., & Fageda, X. (2008). Getting there fast: Globalization, intercontinental flights and location of headquarters. *Journal of Economic Geography*, *8*(4), 471–495. https://doi.org/10.1093/jeg/lbn017

Block, J., & Sandner, P. (2009). What is the effect of the financial crisis on venture capital financing? Empirical evidence from US Internet start-ups. *Venture Capital*, *11*(4), 295–309. https://doi.org/10.1080/13691060903184803

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*, *4*(2), 129–146. https://doi.org/10.1007/s11365-007-0058-8

Bourdeau-Lepage, L., Huriot, J.-M., & Perreur, J. (2009). In Search of Lost Centrality. *Journal of Regional & Urban Economy* , *3*, 549–572. https://doi.org/10.3917/reru.093.0549

Burger, M., & Meijers, E. (2012). Form Follows Function? Linking Morphological and Functional Polycentricity. *Urban Studies*, *49*(5), 1127–1149. https://doi.org/10.1177/0042098011407095

Cassia, L., & Colombelli, A. (2008). Do universities knowledge spillovers impact on new firm's growth? Empirical evidence from UK. *International Entrepreneurship and Management Journal*, *4*(4), 453–465. https://doi.org/10.1007/s11365-008-0084-1

Catalini, C., Guzman, J., & Stern, S. (2019). Hidden in Plain Sight: Venture Growth with or without Venture Capital. *National Bureau of Economic Research.* https://doi.org/10.3386/w26521

Cavallo, A., Ghezzi, A., Colombelli, A., & Casali, G. L. (2020). Agglomeration dynamics of innovative start-ups in Italy beyond the industrial district era. *International Entrepreneurship and Management Journal*, *16*(1), 239–262. https://doi.org/10.1007/s11365-018-0521-8

Centraal Bureau voor de Statistiek (CBS). (n.d.). De Randstad. *Https://Www.Cbs.Nl/Nl-Nl/Nieuws/2022/32/Trek-Uit-de-Randstad-Blijft-Toenemen/de-Randstad.*

Centraal Bureau voor de statistiek (CBS). (2023). Statline: Regionale Kerncijfers Nederland. *Https://Opendata.Cbs.Nl/Statline/#/CBS/Nl/Dataset/70072NED/Table?Fromstatweb.*

Chen, H., Gompers, P., Kovner, A., & Lerner, J. (2010). Buy local? The geography of venture capital. *Journal of Urban Economics*, *67*(1), 90–102. https://doi.org/10.1016/j.jue.2009.09.013

Ciccone, A., & Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, *86*, 54–70.

Combes, P.-P. (2000). Economic Structure and Local Growth: France, 1984–1993. *Journal of Urban Economics*, *47*(3), 329–355. https://doi.org/10.1006/juec.1999.2143

Cristaller, W. (1966). Central places in southern Germany. Englewood Cliffs, N.J.: Presentice-Hall. Translated by C. W. Baskin from Die zentralen orte in Süd-deutchsland, 1933.

Dahl, M. S., & Sorenson, O. (2012). Home sweet home: Entrepreneurs' location choices and the performance of their ventures. *Management Science*, *58*(6), 1059–1071. https://doi.org/10.1287/mnsc.1110.1476

Dalle, J. M., den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research.

den Dulk, C. J., van de Stadt, H., & Vliegen, J. M. (1992). A new measure for degree of urbanization: the address density of the surrounding area. *Maandstatistiek van de Bevolking*, *40*(7), 14–27.

Devine, P. (2003). The A-Z of Social Research: A Dictionary of Key Social Science Research Concepts. SAGE Publications Ltd.

Duranton, G., & Puga, D. (2000). Diversity and Specialisation in Cities: Why, Where and When Does it Matter? In *Urban Studies 37*(3), 533-555.

Duranton, G., & Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies (Vol. 4, pp. 2063-2117). *Handbook of regional and urban economics.*

Eaton, B. C., & Lipsey, R. G. (1982). An Economic Theory of Central Places. In *The Economic Journal*, *92(*365), 56-72. https://www.jstor.org/stable/2232256

Ethical Principles of Psychologists and Code of Conduct. (2017). American Psychological Association (APA).
Etikan, I., Sulaiman Abubakar Musa, & Rukayya Sunusi Alkassim. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11
Eveleens, C. P., van Rijnsoever, F. J., & Niesten, E. M. M. I. (2017). How network-based incubation helps start-up performance: a systematic review against the background of management theories. *Journal of Technology Transfer*, *42*(3), 676–713. https://doi.org/10.1007/s10961-016-9510-7
Felsenstein, D., & Fleischer, A. (2002). Small-scale entrepreneurship and access to capital in peripheral locations: An empirical analysis. *Growth and Change*, *33*(2), 196–215. https://doi.org/10.1111/0017-4815.00187
Florida, R., & Mellander, C. (2016). Rise of the startup city: The changing geography of the venture capital financed innovation. In *California Management Review*, *59(*1), 14-38. https://doi.org/10.1177/0008125616683952
Frenkel, A. (2012). High-Tech Firms' Location Considerations within the Metropolitan Regions and the Impact of Their Development Stages. *European Planning Studies*, *20*(2), 231–255. https://doi.org/10.1080/09654313.2012.651799
Fritsch, M., & Schilder, D. (2008). Does venture capital investment really require spatial proximity? An empirical investigation. *Environment and Planning A*, *40*(9), 2114–2131. https://doi.org/10.1068/a39353
Fritsch, M., & Schilder, D. (2012). The regional supply of venture capital: Can syndication overcome bottlenecks? *Economic Geography*, *88*(1), 59–76. https://doi.org/10.1111/j.1944-8287.2011.01139.x
Gallagher, C., & Miller, P. (1991). New fast-growing companies create jobs. *Long Range Planning*, *24*(1), 96–101. https://doi.org/10.1016/0024-6301(91)90029-N
Ghio, N., Guerini, M., Lehmann, E. E., & Rossi-Lamastra, C. (2015). The emergence of the knowledge spillover theory of entrepreneurship. *Small Business Economics*, *44*(1), 1–18. https://doi.org/10.1007/s11187-014-9588-y
Glaeser, E., Kallal, L., Scheinkman, H. D., & Shleifer, A. (1992). Growth in cities. *Journal of political economy*, *100(*6), 1126-1152.
Goodall, B. (2013). The Economics of Urban Areas (Vol. 3). Pergamon Press.
Guzman, J. (2019). Go west young firm: agglomeration and embeddedness in startup migrations to Silicon Valley. *Columbia Business School Research Paper*, 18–49.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). Multivariate Data Analysis (8th ed.). Cengage Learning.
Hansen, G. S., & Wernerfelt, B. (1989). Determinants of firm performance: The relative importance of economic and organizational factors. *Strategic Management Journal*, *10*(5), 399–411. https://doi.org/10.1002/smj.4250100502
Isenberg, D. J. (2010). How to start an entrepreneurial revolution. *Harvard Business Review*, *88*(6), 40–50.
J. Henderson. (2002). Building the rural economy with high-growth entrepreneurs. *Economic Review-Federal Reserve Bank of Kansas City*, *87*(3), 45-75.
Jacobs, J. (1969). The Economy of Cities. Vintage.
Jarvis, R., Curran, J., Kitching, J., & Lightfoot, G. (2000). The use of quantitative and qualitative criteria in the measurement of performance in small firms, *7*(2), 123-134.
Jeong, J., Kim, J., Son, H., & Nam, D. Il. (2020). The role of venture capital investment in startups' sustainable growth and performance: Focusing on absorptive capacity and venture capitalists' reputation. *Sustainability*, *12*(8), 3447. https://doi.org/10.3390/SU12083447
Kalantaridis, C. (2009). SME Strategy, Embeddedness and Performance in East Cleveland, North East England. *International Small Business Journal*, *27*(4), 496–521.
Kaminski, J., Hopp, C., & Tykvová, T. (2019). New technology assessment in entrepreneurial financing – Does crowdfunding predict venture capital investments? *Technological Forecasting and Social Change*, *139*, 287–302. https://doi.org/10.1016/j.techfore.2018.11.015
Keeble, D., & Walker, S. (1994). New Firms, Small Firms and Dead Firms: Spatial Patterns and Determinants in the United Kingdom. *Regional Studies*, *28*(4), 411–427. https://doi.org/10.1080/00343409412331348366
Li, M., Goetz, S. J., Partridge, M., & Fleming, D. A. (2016). Location determinants of high-growth firms. *Entrepreneurship and Regional Development*, *28*(1–2), 97–125. https://doi.org/10.1080/08985626.2015.1109003
Marshall, A. (1920). Principles of Economics Macmillan. *London*, *8th ed.*
McEwan, B. (2020). Sampling and validity. In *Annals of the International Communication Association*, *44*(3), 235-247. Routledge. https://doi.org/10.1080/23808985.2020.1792793
Meccheri, N., & Pelloni, G. (2006). Rural entrepreneurs and institutional assistance: An empirical study from mountainous Italy. *Entrepreneurship and Regional Development*, *18*(5), 371–392. https://doi.org/10.1080/08985620600842113
Melissa J. Goertzen. (2017). Introduction to Quantitative Research and Data. In *Applying Quantitative Methods to E-Book Collections*, *53*(4), 12–18. Library Technology Reports.

Minai, M. S., & Lucky, E. O. I. (2011). The Moderating Effect of Location on Small Firm Performance: Empirical Evidence. *International Journal of Business and Management*, *6*(10), 178. https://doi.org/10.5539/ijbm.v6n10p178

Mulligan, G. F. (1984). Agglomeration and central place theory: A review of the literature. *International Regional Science Review*, *9*(1), 1–42.

Murphy, G. B., Trailer, J. W., & Hill, R. C. (1996). Measuring Performance in Entrepreneurship Research. *Journal of Business Research*, *36*(1), 15–23.

Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J., & Olander, L.-O. (2011). The Dynamics of Agglomeration Externalities along the Life Cycle of Industries. *Regional Studies*, *45*(1), 49–65. https://doi.org/10.1080/00343401003596307

Oakey, R. (2007). Clustering and the R&D management of high-technology small firms: In theory and practice. *R and D Management*, *37*(3), 237–248. https://doi.org/10.1111/j.1467-9310.2007.00472.x

Oerlemans, L. A. G., & Meeus, M. T. H. (2005). Do organizational and spatial proximity impact on firm performance? *Regional Studies*, *39*(1), 89–104. https://doi.org/10.1080/0034340052000320896

Pan, F., & Yang, B. (2019). Financial development and the geographies of startup cities: evidence from China. *Small Business Economics*, *52*(3), 743–758. https://doi.org/10.1007/s11187-017-9983-2

Parr, J. (2004). The Polycentric Urban Region: A Closer Inspection. *Regional Studies*, *38*(3), 231–240. https://doi.org/10.1080/003434042000211114

Parr, J. B. (2017). Central Place Theory: An Evaluation. *Review of Urban and Regional Development Studies*, *29*(3), 151–164. https://doi.org/10.1111/rurd.12066

Porter, M. E. (1998). The Adam Smith address: Location, clusters, and the “new” microeconomics of competition. *Business Economics*, *33*(1), 7–13.

Qian, H., Haynes, K. E., & Riggle, J. D. (2011). Incubation push or business pull? investigating the geography of U.S. Business incubators. *Economic Development Quarterly*, *25*(1), 79–90. https://doi.org/10.1177/0891242410383275

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2119–2171). Elsevier. https://doi.org/10.1016/S1574-0080(04)80006-3

Schutjens, V. A. J. M., & Wever, E. (2005). Determinants of new firm success. *Papers in Regional Science*, *79*(2), 135–159. https://doi.org/10.1111/j.1435-5597.2000.tb00765.x

Stam, E. (2005). The Geography of Gazelles in the Netherlands. *Tijdschrift Voor Economische En Sociale Geografie*, *96*(1), 121–127. https://doi.org/10.1111/j.1467-9663.2005.00443.x

Stam, E. (2014). *The Dutch Entrepreneurial Ecosystem*.

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370. https://doi.org/10.1093/jnlecg/lbh027

Thwaites, A., & Wynarczyk, P. (1996). The economic performance of innovative small firms in the South East region and elsewhere in the UK. *Regional Studies*, *30*(2), 135–149. https://doi.org/10.1080/00343409612331349538

Tunberg, M. (2014). Approaching rural firm growth: A literature review. *Journal of Enterprising Communities*, *8*(4), 261–286. https://doi.org/10.1108/JEC-11-2013-0039

van Geenhuizen, M., & Reyes-Gonzalez, L. (2007). Does a clustered location matter for high-technology companies' performance? The case of biotechnology in the Netherlands. *Technological Forecasting and Social Change*, *74*(9), 1681–1696. https://doi.org/10.1016/j.techfore.2006.10.009

Van Rijnsoever, F. J., Van Weele, M. A., & Eveleens, C. P. (2017). Network brokers or hit makers? Analyzing the influence of incubation on start-up investments. *International Entrepreneurship and Management Journal*, *13*(2), 605–629. https://doi.org/10.1007/s11365-016-0416-5

Wennberg, K., Wiklund, J., & Wright, M. (2011). The effectiveness of university knowledge spillovers: Performance differences between university spinoffs and corporate spinoffs. *Research Policy*, *40*(8), 1128–1143. https://doi.org/10.1016/j.respol.2011.05.014

Yu, L., Orazem, P. F., & Jolly, R. W. (2011). Why do rural firms live longer? *American Journal of Agricultural Economics*, *93*(3), 669–688. https://doi.org/10.1093/ajae/aaq173

Zook, M. A. (2002). Grounded capital: venture financing and the geography of the Internet industry. *Journal of Economic Geography*, *2*(2), 151–177.

APPENDICES

APPENDIX 1: INITIAL MULTICOLLINEARITY PROBLEMS

In the initial stage of this research, the methodological approach hinged on evaluating agglomeration economies through the lens of address density as a continuous variable. However, this approach ran aground due to issues of severe multicollinearity. Specifically, the Variance Inflation Factor (VIF) values for the interaction effects spiked to around 30 (see table appendix 1a), thus severely undermining the reliability of interpreting these interaction effects. Further details on these initial multicollinearity diagnostics can be found in the table below.

The path to resolution was laden with complexities. The investigation into potential remedies was hindered by a series of failed attempts with multiple indicators. The difficulty stemmed from the inherent nature of this research - the independent variables are conceptually intertwined, thereby compounding the multicollinearity issue.

However, after a painstaking exploration of various possibilities, a breakthrough was achieved. It was discerned that adopting the five-level-classification of urban density, as proposed by den Dulk et al. (1992) and also adopted by the Central Bureau of Statistics of the Netherlands (CBS), significantly alleviated the multicollinearity. Despite that our measure is not continuous anymore, the scatterplots are still presented as the values and relation can still be valuable for comprehending the relationship

Table Appendix 1a. Initial Collinearity Diagnostics

Variable	Tolerance	VIF [a]
Number of Employees	,817	1,224
Number of Funding Rounds	,819	1,221
Address density (per km2)	,481	2,080
Intermediate Zone	,196	5,111
National Periphery	,035	28,895
Adress density * Intermediate Zone	,183	5,462
Address density * National Periphery	,034	29,242

VIF = Variation Inflation Factor, calculated by 1/Tolerance.

APPENDIX 2: CLASSIFICATION DETAILS OF INDEPENDENT PREDICTORS

Table Appendix 2a. Classification Geographic Centrality in the Netherlands (CBS, n.d.)

Central zone	Covering the provinces of Utrecht and Zuid-Holland and the southern part of Noord-Holland)
Intermediate zone	Covering Flevoland, Noord-Brabant, Gelderland (excluding the Achterhoek), and the Noord-Holland regions of IJmond and Alkmaar
National periphery	Covering the three northern provinces of Groningen, Friesland, and Drenthe, as well as Zeeland and Limburg, and a large part of Overijssel, the Achterhoek, and the Kop van Noord-Holland

Table Appendix 2b. Address Densities (per km^2) of the Municipalities (CBS, 2023)

Aa en Hunze	41
Aalsmeer	665
Aalten	125
Achtkarspelen	120
Alblasserdam	959
Albrandswaard	490
Alkmaar	480
Almelo	505
Almere	688
Alphen aan den Rijn	392
Alphen-Chaam	48
Altena	114
Ameland	31
Amersfoort	1120
Amstelveen	1031
Amsterdam	2771
Apeldoorn	220
Arnhem	809
Assen	399
Asten	102
Baarle-Nassau	39
Baarn	362
Barendrecht	1001
Barneveld	133
Beek (L.)	356
Beekdaelen	214
Beesel	218
Berg en Dal	184
Bergeijk	81
Bergen (L.)	56
Bergen (NH.)	155
Bergen op Zoom	390
Berkelland	76
Bernheze	144
Best	389
Beuningen	260
Beverwijk	1070
De Bilt	299
Bladel	119
Blaricum	481
Bloemendaal	256
Bodegraven-Reeuwijk	197
Boekel	130
Borger-Odoorn	41
Borne	396
Borsele	70
Boxtel	210
Breda	683
Brielle	299
Bronckhorst	58
Brummen	114
Brunssum	842
Bunnik	178
Bunschoten	287
Buren	82
Capelle aan den IJssel	2245
Castricum	328
Coevorden	53
Cranendonck	119
Culemborg	437
Dalfsen	74
Dantumadiel	97
Delft	2300
Deurne	121
Deventer	356
Diemen	1345
Dijk en Waard	592
Dinkelland	63
Doesburg	461
Doetinchem	338
Dongen	395
Dordrecht	725
Drechterland	144
Drimmelen	126
Dronten	54
Druten	222
Duiven	318
Echt-Susteren	143
Edam-Volendam	287
Ede	159
Eemnes	132
Eemsdelta	86
Eersel	104
Eijsden-Margraten	147
Eindhoven	1319
Elburg	153
Emmen	149
Enkhuizen	700
Enschede	542
Epe	95
Ermelo	134
Etten-Leur	348
De Fryske Marren	67
Geertruidenberg	370
Geldrop-Mierlo	584
Gemert-Bakel	105
Gennep	159
Gilze en Rijen	174
Goeree-Overflakkee	85
Goes	204
Goirle	243
Gooise Meren	652
Gorinchem	938
Gouda	2065
's-Gravenhage (gemeente)	3246
Groningen (gemeente)	651
Gulpen-Wittem	94
Haaksbergen	101
Haarlem	2653
Haarlemmermeer	334
Halderberge	182
Hardenberg	81
Harderwijk	543
Hardinxveld-Giessendam	446
Harlingen	314
Hattem	234
Heemskerk	648
Heemstede	1398
Heerde	104
Heerenveen	126
Heerlen	1030
Heeze-Leende	68
Heiloo	593
Den Helder	630
Hellendoorn	109
Hellevoetsluis	441
Helmond	788
Hendrik-Ido-Ambacht	1238
Hengelo (O.)	645
's-Hertogenbosch	678
Heumen	184
Heusden	248
Hillegom	777
Hilvarenbeek	73
Hilversum	952
Hoeksche Waard	144
Hof van Twente	72
Het Hogeland	48
Hollands Kroon	59
Hoogeveen	201
Hoorn	1668
Horst aan de Maas	96
Houten	380
Huizen	1207
Hulst	68
IJsselstein	689
Kaag en Braassem	192
Kampen	162
Kapelle	148
Katwijk	1097
Kerkrade	1095
Koggenland	120
Krimpen aan den IJssel	1639
Krimpenerwaard	163
Laarbeek	178
Land van Cuijk	115
Landgraaf	738
Landsmeer	218

Lansingerland	466
Laren (NH.)	439
Leeuwarden	269
Leiden	2819
Leiderdorp	1089
Leidschendam-Voorburg	1141
Lelystad	151
Leudal	100
Leusden	228
Lingewaard	326
Lisse	664
Lochem	72
Loon op Zand	210
Lopik	77
Losser	101
Maasdriel	155
Maasgouw	242
Maashorst	188
Maassluis	1892
Maastricht	1134
Medemblik	160
Meerssen	325
Meierijstad	192
Meppel	294
Middelburg (Z.)	497
Midden-Delfland	171
Midden-Drenthe	43
Midden-Groningen	102
Moerdijk	105
Molenlanden	97
Montferland	152
Montfoort	154
Mook en Middelaar	212
Neder-Betuwe	150
Nederweert	76
Nieuwegein	1285
Nieuwkoop	150
Nijkerk	256
Nijmegen	1583
Nissewaard	545
Noardeast-Fryslân	55
Noord-Beveland	53
Noordenveld	75
Noordoostpolder	45
Noordwijk	353
Nuenen, Gerwen en Nederwetten	313
Nunspeet	88
Oegstgeest	1518
Oirschot	81
Oisterwijk	178
Oldambt	83
Oldebroek	97
Oldenzaal	684
Olst-Wijhe	70
Ommen	43
Oost Gelre	118
Oosterhout	358
Ooststellingwerf	51
Oostzaan	358
Opmeer	121
Opsterland	59
Oss	255
Oude IJsselstreek	129
Ouder-Amstel	266
Oudewater	111
Overbetuwe	186
Papendrecht	1545
Peel en Maas	118
Pekela	118
Pijnacker-Nootdorp	602
Purmerend	449
Putten	118
Raalte	95
Reimerswaal	92
Renkum	328
Renswoude	119
Reusel-De Mierden	72
Rheden	260
Rhenen	197
Ridderkerk	915
Rijssen-Holten	159
Rijswijk (ZH.)	1998
Roerdalen	109
Roermond	479
De Ronde Venen	188
Roosendaal	337
Rotterdam	1461
Rozendaal	25
Rucphen	155
Schagen	124
Scherpenzeel	307
Schiedam	2120
Schiermonnikoog	14
Schouwen-Duiveland	74
Simpelveld	325
Sint-Michielsgestel	213
Sittard-Geleen	594
Sliedrecht	871
Sluis	51
Smallingerland	222
Soest	456
Someren	104
Son en Breugel	291
Stadskanaal	129
Staphorst	45
Stede Broec	642
Steenbergen	72
Steenwijkerland	71
Stein (L.)	564
Stichtse Vecht	300
Súdwest-Fryslân	82
Terneuzen	109
Terschelling	26
Texel	43
Teylingen	579
Tholen	76
Tiel	562
Tilburg	826
Tubbergen	59
Twenterand	131
Tynaarlo	103
Tytsjerksteradiel	95
Uitgeest	299
Uithoorn	746
Urk	522
Utrecht (gemeente)	1732
Utrechtse Heuvelrug	167
Vaals	245
Valkenburg aan de Geul	228
Valkenswaard	273
Veendam	171
Veenendaal	1498
Veere	83
Veldhoven	640
Velsen	703
Venlo	388
Venray	120
Vijfheerenlanden	168
Vlaardingen	1532
Vlieland	14
Vlissingen	672
Voerendaal	185
Voorschoten	1038
Voorst	88
Vught	231
Waadhoeke	75
Waalre	357
Waalwijk	347
Waddinxveen	470
Wageningen	603
Wassenaar	240
Waterland	143
Weert	224
Weesp	419
West Betuwe	98
West Maas en Waal	111
Westerkwartier	77
Westerveld	31
Westervoort	963
Westerwolde	43
Westland	574
Weststellingwerf	54
Westvoorne	132
Wierden	106
Wijchen	276
Wijdemeren	228
Wijk bij Duurstede	217
Winterswijk	96
Woensdrecht	109
Woerden	255
De Wolden	47
Wormerland	183
Woudenberg	148
Zaanstad	948
Zaltbommel	152
Zandvoort	303
Zeewolde	37
Zeist	615
Zevenaar	222
Zoetermeer	1646
Zoeterwoude	186
Zuidplas	322
Zundert	81
Zutphen	563
Zwartewaterland	110
Zwijndrecht	1022
Zwolle	544

APPENDIX 3: MULTICOLLINEARITY DIAGNOSTICS WITH NEW MEASURE

The dependent variables, Total Funding and Total Revenue, exhibit strong correlations with the variables (see table appendix 3a). These significant and robust relationships suggest a potential pronounced influence of these factors on the financial outcomes of startups. Urban Density shows significant correlations with all other variables. Particularly, it demonstrates a strong correlation with Centrality, indicated by a robust Spearman's rank correlation coefficient of 0,628. This strong correlation implies that Centrality and Urban Density tend to change together, which may introduce multicollinearity into the models. Therefore, it is crucial to closely examine the potential multicollinearity between Centrality and Urban Density.

Multicollinearity, characterized by high correlations between independent variables, can inflate the variances of the parameter estimates and render these estimates unreliable (Hair et al., 2018). Consequently, while the models may explain the relationships within the data, caution is necessary when interpreting the individual predictors to avoid potential misleading conclusions.

Table Appendix 3a: Pairwise comparisons

Variable [a]	**1**	**2**	**3**	**4**	**5**	**6**
1. Total Funding (in million dollars)	-	-	-	-	-	
2. Total Revenue (in million dollars)	-	-	-	-	-	
3. Number of Employees	0,729**	0,653**	-	-	-	
4. Number of Investors	0,568**	0,334**	0,419**	-	-	
5. Number of Funding Rounds	0,556**	0,321**	0,392**	0,699**	-	
6. Urban density [b]	0,205**	0,222**	0,159**	0,202**	0,157**	
7. Centrality [c]:	0,113**	0,162**	0,106**	0,065	0,077*	0,628**

[a] N = 727 startups; [bc] Spearman’s rank used for correlation; **p ≤ 0,01, * p ≤ 0,05

Given these concerns, and standard assumption checking, an in-depth analysis of multicollinearity diagnostics is conducted (see table appendix 3b). The diagnostics for multicollinearity within the data presented an intricate landscape. In assessing the control variables, a minimum tolerance of 0,811 and a Variance Inflation Factor (VIF) of 1,23 was observed. However, the direct effects’ lowest tolerance level is 0,172 and a corresponding VIF of 5,80. This marginally crosses the acceptable threshold for multicollinearity, prompting a need for caution in interpreting the interaction terms.

The scenario further complexified upon examining the interaction effects. Notably, three variables were excluded from the analysis due to a lack of intersecting data. This exclusion was justified, as depicted by the zero values in certain cells in table appendix 3b. Additionally, the statistical software made the decision to exclude seven variables. As a result of these variables contributing zero value to the model which is corroborated by their corresponding tolerance values of zero and non-applicable VIF values, given that VIF is the reciprocal of tolerance (Hair et al., 2018). Despite these exclusions, five interaction terms are left possessing acceptable VIF and tolerance levels. However, given the number of variables eliminated from the analysis, careful interpretation is required.

Table Appendix 3b. Multicollinearity diagnostics

Variable	Tolerance	VIF [a]
(Log) Number of Employees	0,811	1,233
(Log) Number of Investors	0,811	1,233
Not Urbanized	0,451	2,22
Hardly Urbanized	0,186	5,37
Moderately Urbanized	0,558	1,792
Strongly Urbanized	0,796	1,257
Intermediate Zone	0,3	3,338
National Periphery	0,172	5,8
Not Urbanized * National Periphery	0,338	2,955
Hardly Urbanized* National Periphery	0,00	N/A
Moderately Urbanized * National Periphery	0,515	1,943
Strongly Urbanized * National Periphery	-	-
Extremely Urbanized * National Periphery	-	-
Not Urbanized * Intermediate Zone	0,442	2,261
Hardly Urbanized * Intermediate Zone	0,00	N/A
Moderately Urbanized * Intermediate Zone	-	-
Strongly Urbanized * Intermediate Zone	0,578	1,731
Extremely Urbanized * Intermediate Zone	0,00	N/A
Not Urbanized * Central Zone	0,00	N/A
Hardly Urbanized * Central Zone	0,493	2,028
Moderately Urbanized * Central Zone	0,00	N/A
Strongly Urbanized * Central Zone	0,00	N/A
Extremely Urbanized * Central Zone	0,00	N/A

VIF = Variation Inflation Factor, calculated by 1/Tolerance

APPENDIX 4: LINEARITY DIAGNOSTICS WITH TOTAL FUNDING

In this appendix the linearity diagnostics are displayed. Although the continues variable of Address Density and the Number of Investors do not exist in the regression model of this research, they are displayed in sake of transparency, and possibly helping to understand the relationships of these variables on the dependent variables' indicators.

Linearity Diagnostics with Total Funding

Across Appendices 4a-4d, different degrees of relationships between Address Density, Number of Employees, Number of Investors, Number of Funding Rounds, and Total Funding are observed in the scatterplots. Appendix 4a shows Address Density having a marginal positive effect, explaining only 0,55% of the variance. In contrast, Number of Employees (Appendix 4b) exhibits a strong positive correlation, accounting for a significant 53,1% of the variance. Appendices 4c and 4d, representing Number of Investors and Number of Funding Rounds respectively, display moderate positive effects on Total Funding. Although the polynomials in all scatterplots are statistically significant, their contribution to the explained variance is minimal. This suggests that linear regression models provide more interpretable results for all the examined factors.

Linearity Diagnostics with Total Revenue

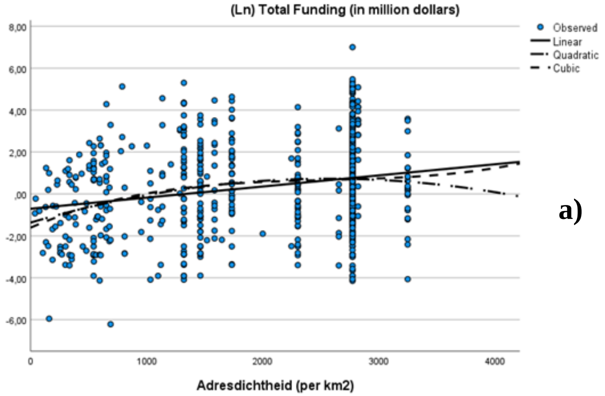
Appendix 4e (Address Density) reveals a weak, yet significant, positive relationship with Total Revenue, accounting for 0,046 of the variance. However, despite significant polynomials, their marginal contribution indicates a linear model as more suitable for interpretation. Appendix 4f (Number of Employees) exhibits a strong positive correlation, explaining 0,426 of the variance, with significant polynomials enhancing the R-square by 6,6%. Despite the significant improvement the linear function is used in sake of interpretability and the low added value of another transformation of an already transformed control variable (Hair et al., 2018). Appendices 4g (Number of Investors) and 4h (Number of Funding Rounds), both display weak to moderate positive correlations, explaining 0,111 and 0,103 of the variance respectively. Despite significant polynomials, their minimal R-square contribution favors linear models. The findings indicate a model choice depends on the relationship's strength and complexity, maintaining predictive power while ensuring interpretability.

Model Summary and Parameter Estimates

Dependent Variable: (Ln) Total Funding (in million dollars)

Equation	Model Summary: R Square	F	df1	df2	Sig.	Parameter Estimates: Constant	b1	b2	b3
Linear	,052	37,894	1	696	<,001	-,705	,001		
Quadratic	,060	22,275	2	695	<,001	-1,366	,002	-3,165E-7	
Cubic	,061	15,020	3	694	<,001	-1,607	,002	-8,917E-7	1,171E-10

The independent variable is Adresdichtheid (per km2).



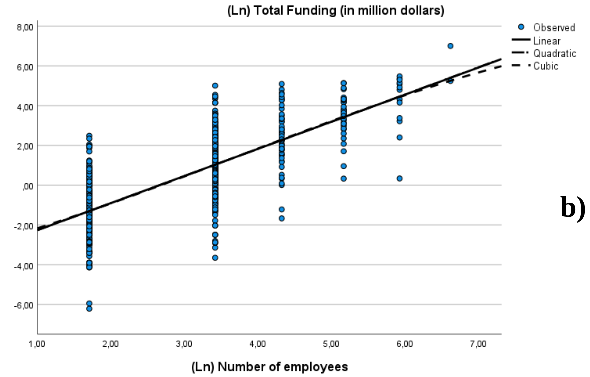
a) Address Density (per km^2)

Model Summary and Parameter Estimates

Dependent Variable: (Ln) Total Funding (in million dollars)

Equation	Model Summary: R Square	F	df1	df2	Sig.	Parameter Estimates: Constant	b1	b2	b3
Linear	,531	788,075	1	696	<,001	-3,628	1,363		
Quadratic	,531	393,472	2	695	<,001	-3,620	1,357	,001	
Cubic	,531	261,984	3	694	<,001	-3,241	,965	,120	-,011

The independent variable is (Ln) Number of employees.



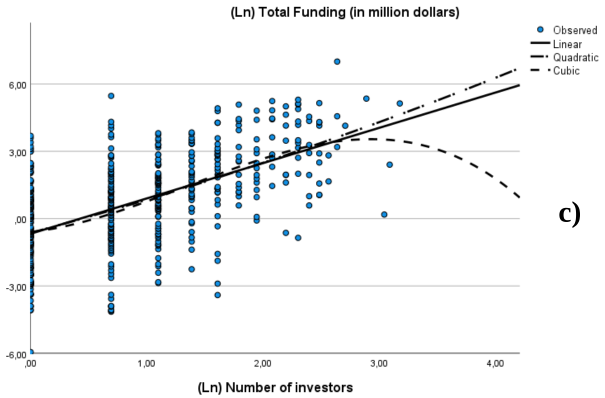
b) Number of Employees

Model Summary and Parameter Estimates

Dependent Variable: (Ln) Total Funding (in million dollars)

Equation	Model Summary: R Square	F	df1	df2	Sig.	Parameter Estimates: Constant	b1	b2	b3
Linear	,323	261,627	1	548	<,001	-,670	1,572		
Quadratic	,324	130,927	2	547	<,001	-,633	1,396	,083	
Cubic	,327	88,457	3	546	<,001	-,591	,563	1,060	-,263

The independent variable is (Ln) Number of investors.



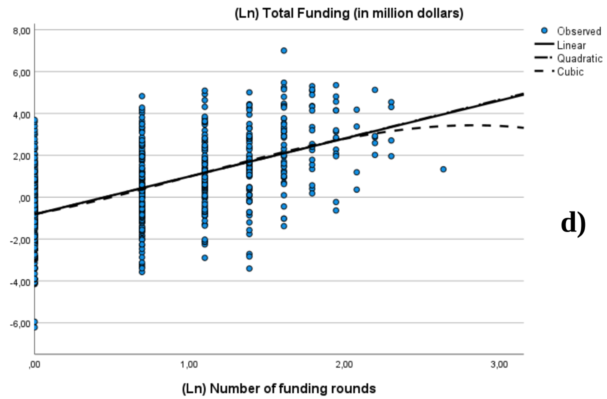
c) Number of Investors

Model Summary and Parameter Estimates

Dependent Variable: (Ln) Total Funding (in million dollars)

Equation	Model Summary: R Square	F	df1	df2	Sig.	Parameter Estimates: Constant	b1	b2	b3
Linear	,311	314,685	1	696	<,001	-,817	1,811		
Quadratic	,311	157,119	2	695	<,001	-,815	1,795	,010	
Cubic	,312	104,834	3	694	<,001	-,805	1,378	,612	-,201

The independent variable is (Ln) Number of funding rounds.



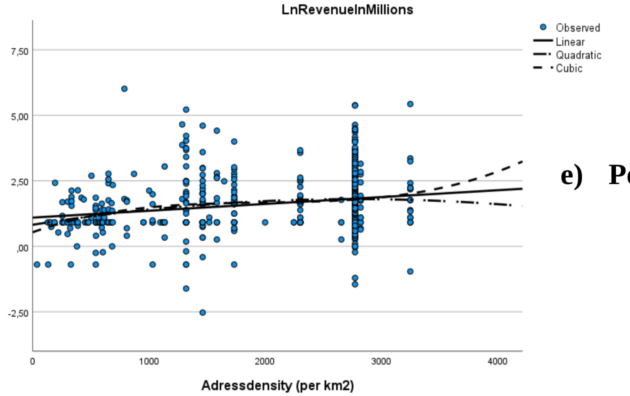
d) Number of Funding Rounds

Model Summary and Parameter Estimates

Dependent Variable: LnRevenueInMillions

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	,046	27,350	1	563	<,001	1,097	,000		
Quadratic	,051	15,205	2	562	<,001	,824	,001	-1,279E-7	
Cubic	,054	10,774	3	561	<,001	,539	,002	-7,674E-7	1,290E-10

The independent variable is Adressdensity (per km2).



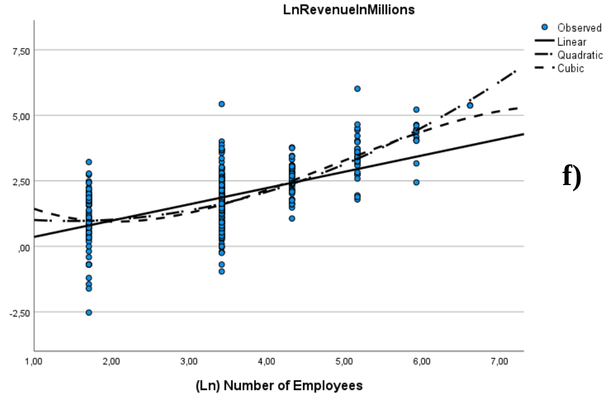
e) Population Density (per km^2)

Model Summary and Parameter Estimates

Dependent Variable: LnRevenueInMillions

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	,426	417,467	1	563	<,001	-,259	,621		
Quadratic	,492	271,657	2	562	<,001	1,345	-,511	,174	
Cubic	,495	183,535	3	561	<,001	3,024	-2,238	,694	-,047

The independent variable is (Ln) Number of Employees.



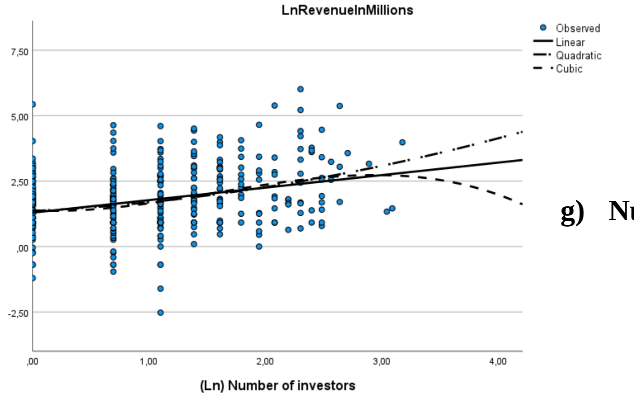
f) Number of Employees

Model Summary and Parameter Estimates

Dependent Variable: LnRevenueInMillions

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	,111	57,013	1	455	<,001	1,297	,478		
Quadratic	,116	29,856	2	454	<,001	1,358	,217	,119	
Cubic	,119	20,485	3	453	<,001	1,382	-,203	,604	-,129

The independent variable is (Ln) Number of investors.



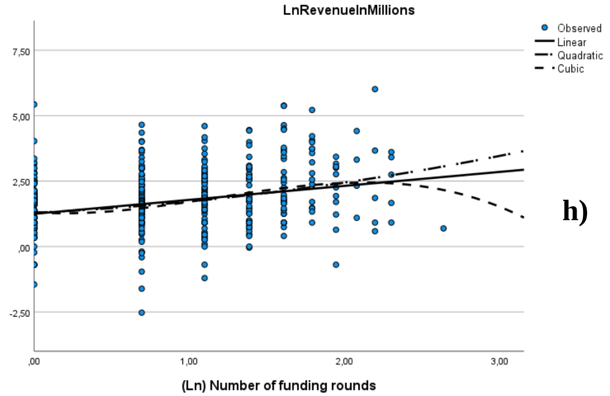
g) Number of Investors

Model Summary and Parameter Estimates

Dependent Variable: LnRevenueInMillions

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	,103	64,489	1	563	<,001	1,252	,533		
Quadratic	,106	33,286	2	562	<,001	1,290	,285	,147	
Cubic	,111	23,454	3	561	<,001	1,313	-,411	1,133	-,324

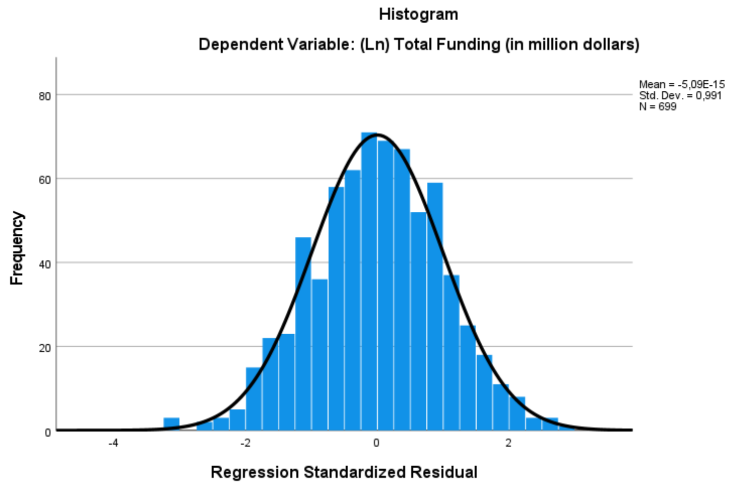
The independent variable is (Ln) Number of funding rounds.



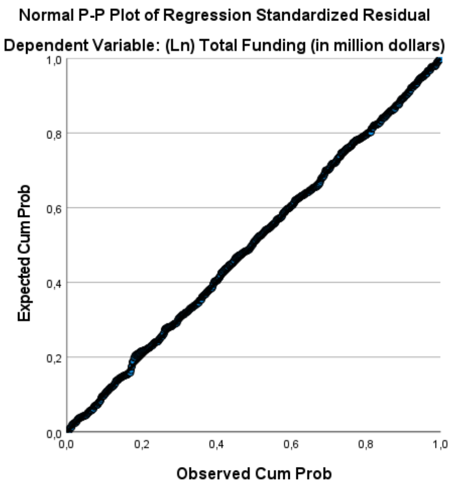
h) Number of Funding Rounds

APPENDIX 5: NORMALITY OF RESIDUALS CHECKS

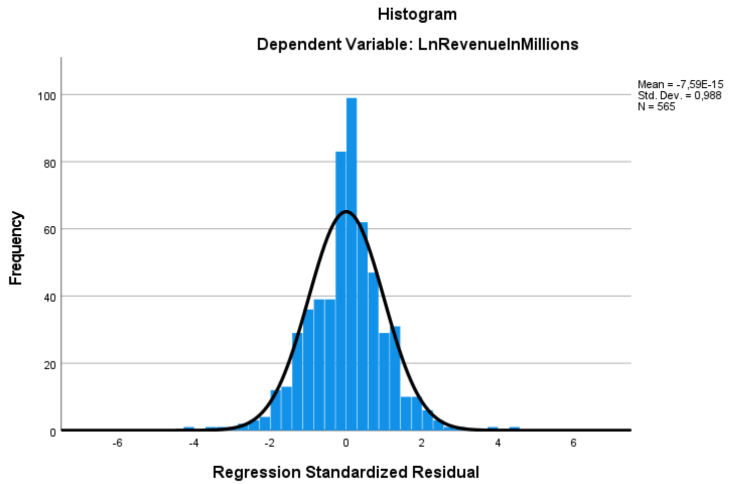
a. Histogram of Residuals (Total Funding)



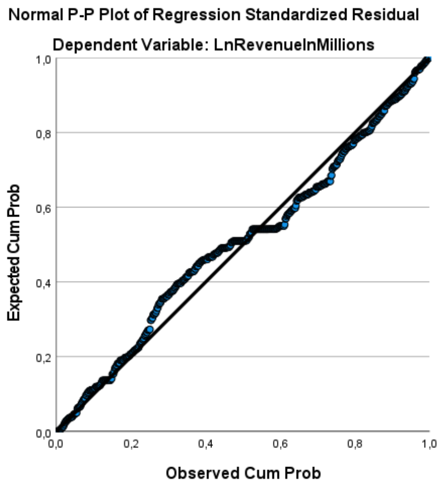
b. Normal P-P Plot of Residuals (Total Funding)



c. Histogram of Residuals (Total Revenue)



d. Normal P-P Plot of Residuals (Total Funding)

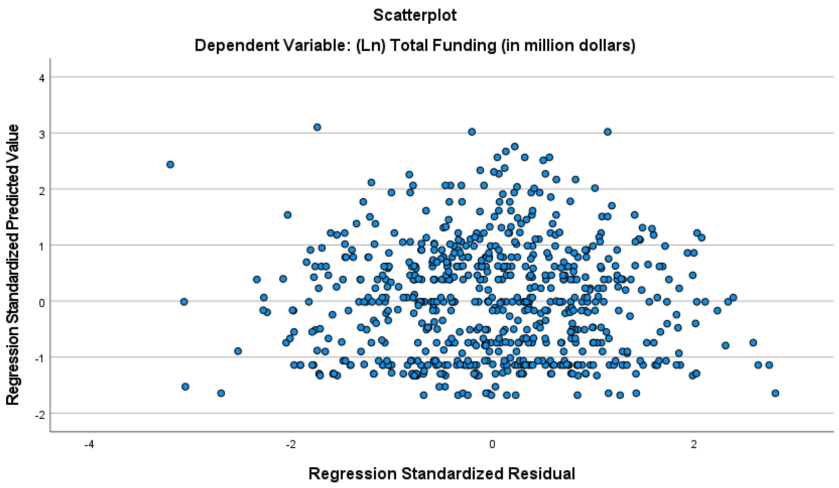


Examination of the histograms and P-P plots presented in Appendices 5a and 5b confirms the normal distribution of error terms for models predicting Total Funding, thereby satisfying the assumption of normality of error terms. This adherence bolsters the validity of these models and strengthens the derived conclusions. In contrast, Appendices 5c and 5d, representing the models predicting Total Revenue, exhibit slight violations of the normality assumption. This deviation warrants cautious interpretation of the respective findings, emphasizing the importance of carefully considering potential skewness or kurtosis in the residuals that may impact the reliability of these results.

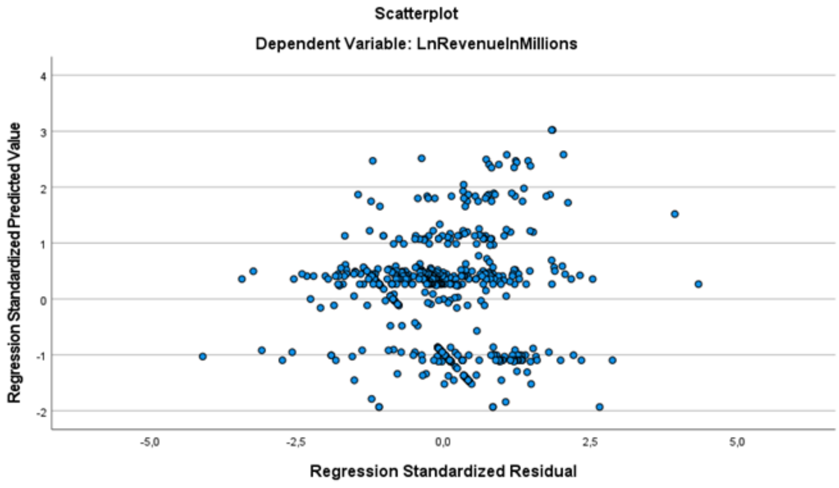
APPENDIX 6: HOMOSCEDASTICITY CHECKS

Upon inspecting the scatterplots in Appendices 6a and 6b, which represent residuals plotted against the predicted values for total funding and total revenue respectively, no distinct patterns are discernible. This lack of patterning in the dispersion of residuals suggests the assumption of homoscedasticity, namely the consistent variance of errors, holds valid for both models.

a) Scatterplot of Residuals (Total Funding)



b) Scatterplot of Residuals (Total Revenue)



APPENDIX 7: DATA AGGREGATION AND INDUSTRY CLASSIFICATION

In this study, a key aspect of the analysis involved aggregating the raw data into distinct industry categories. This process allowed us to examine the relationships between agglomeration economies, geographic centrality, and startup performance within broader industry contexts. To achieve this, the initial provided industry classifications were aggregated into higher-level categories to facilitate a comprehensive analysis. The aim was to strike a balance between capturing industry-specific nuances and ensuring sufficient sample sizes (minimum n = 30) within each category for the statistical analysis.

The resulting industry categories, namely Technology, Services, Healthcare, Retail and Commerce, and Other, encompassed a wide range of sectors and represented broad industry groupings. See table appendix 7a for the subcategories and the number of startups in each overarching industry group.

It is important to note that the classification process inherently involved some subjectivity and interpretation due to the diverse nature of the industries and potential overlaps between sectors. Nevertheless, consistency is ensured by carefully considering the descriptions and characteristics of the individual industry groups. See table appendix 7b for the model 3 results for each industry group.

Table Appendix 7a. Aggregated industries

Overarching Industry (*N*)	**Subcategories**
Technology (302 startups)	Artificial Intelligence
	Apps
	Automation technology
	Biotechnology
	Consumer electronics
	Data and analytics
	Information Technology
	Internet Services
	Messaging and Telecommunications
	Modular intelligence
	Platforms (software)
Services (128 startups)	Administrative Services
	Advertising
	Advising
	Consulting
	Design
	Education
	Financial services
	Professional Services
Health Care (43 startups)	Health Care
Retail and Commerce (77 startups)	Commerce and shopping
	Real Estate
Other (149 startups)	Agriculture
	Clothing
	Consumer goods
	Energy
	Events
	Food and Beverage
	Gaming
	Government and Military
	Hardware
	Manufacturing
	Media and Entertainment
	Science and Engineering
	Transportation
	Travel and Tourism

Table 7b. Model Summary and Regression Coefficients of Robustness Tests across Industries

	Health Care	**Retail and Commerce**	**Services**	**Technology**	**Other**
(Constant)	-5,418 **** (-6,034)	-3,580 **** (-10,576)	-3,411 **** (-8,485)	-3,288 **** (-13,916)	-3,327 *** (-9,620)
(Log) Number of Employees	1,709 **** 6,474	1,036 **** (8,776)	1,061 **** (8,408)	1,129 **** (15,349)	1,014 **** (9,825)
(Log) Number of Funding Rounds	1,037 ** (2,746)	1,058 **** (4,579)	0,892 **** (4,011)	0,913 **** (6,976)	1,237 **** (7,281)
Extremely Urbanized Δ Strongly Urbanized	-0,928 (1,265)	-0,012 (0,028)	-0,699 * (1,664)	-0,253 (1,158)	0,411 (-1,329)
Extremely Urbanized Δ Moderately Urbanized	-0,384 (0,558)	0,614 (-1,483)	0,003 (-0,008)	0,225 (-0,840)	0,272 (-0,832)
Extremely Urbanized Δ Hardly Urbanized	-0,468 (0,418)	0,424 (-,685)	-0,324 (0,530)	0,896 *** (-2,657)	0,315 (-0,658)
Extremely Urbanized Δ Not Urbanized	0,415 (-0,436)	-0,337 (0,603)	-0,111 (0,187)	1,254 **** (-3,614)	0,851 ** (-2,028)
Central Zone Δ Intermediate Zone	0,237 (-0,414)	-0,308 (0,675)	1,642 *** (-2,830)	0,099 (-0,344)	-0,412 (1,149)
Central Zone Δ National Periphery	-0,164 (0,146)	-1,722 ** (2,363)	0,911 (-1,249)	-0,671 ** (2,011)	-0,064 (0,126)
Diagnostics					
R	0,841	0,868	0,784	0,789	0,824
Adjusted R^2	0,638	0,724	0,589	0,612	0,661
N	43	77	128	302	149
Degrees of freedom	8	8	8	8	8
F – value	10,245 ****	25,981 ****	23,791 ****	60,387 ****	1,0136

* p ≤ 0,10, **p ≤ 0,05, ***p ≤ 0,01, ****p ≤ 0,001; (Constant) = (Ln) Total Funding in million dollars; accompanying t-values in parentheses; Δ = in comparison of